High Availability Cluster
Multi-Processing for AIX

Administration Guide

Version 5.4

SC23-4862-09

Ninth Edition (August 2006)

Before using the information in this book, read the general information in Notices for HACMP Administration Guide.

This edition applies to HACMP for AIX 5L, version 5.4 and to all subsequent releases of this product until otherwise indicated in new editions.

Contents

About This Guide

This guide provides information necessary to configure, manage, and troubleshoot the High Availability Cluster Multi-Processing for AIX 5L (HACMP) software.

The following table provides version and manual part numbers for the *Administration Guide*.

HACMP Version	Book Name	Book Number
5.4	*Administration Guide*	SC23-4862-09
5.3 last update 08/2006	*Administration Guide*	SC23-4862-08
5.3	*Administration Guide*	SC23-4862-06
5.2 last update 10/2005	*Administration and Troubleshooting Guide*	SC23-4862-05
5.1	*Administration and Troubleshooting Guide*	SC23-4862-02
4.5	*Administration Guide*	SC23-4279-05

Who Should Use This Guide

This guide is intended for system administrators and customer engineers responsible for configuring, managing, and troubleshooting an HACMP cluster. As a prerequisite for maintaining the HACMP software, you should be familiar with:

- IBM eServer pSeries system components (including disk devices, cabling, and network adapters)
- The AIX 5L operating system, including the Logical Volume Manager subsystem
- The System Management Interface Tool (SMIT)
- Communications, including the TCP/IP subsystem.

Highlighting

This guide uses the following highlighting conventions:

Italic	Identifies new terms or concepts, or indicates emphasis.
Bold	Identifies routines, commands, keywords, files, directories, menu items, and other items whose actual names are predefined by the system.
`Monospace`	Identifies examples of specific data values, examples of text similar to what you might see displayed, examples of program code similar to what you might write as a programmer, messages from the system, or information that you should actually type.

ISO 9000

ISO 9000 registered quality systems were used in the development and manufacturing of this product.

HACMP Publications

The HACMP software comes with the following publications:

- *HACMP for AIX 5L Release Notes* in **/usr/es/sbin/cluster/release_notes** describe issues relevant to HACMP on the AIX platform: latest hardware and software requirements, last-minute information on installation, product usage, and known issues.
- *HACMP on Linux Release Notes* in **/usr/es/sbin/cluster/release_notes.linux/** describe issues relevant to HACMP on the Linux platform: latest hardware and software requirements, last-minute information on installation, product usage, and known issues.
- *HACMP for AIX 5L: Administration Guide,* SC23-4862
- *HACMP for AIX 5L: Concepts and Facilities Guide,* SC23-4864
- *HACMP for AIX 5L: Installation Guide,* SC23-5209
- *HACMP for AIX 5L: Master Glossary,* SC23-4867
- *HACMP for AIX 5L: Planning Guide,* SC23-4861
- *HACMP for AIX 5L: Programming Client Applications*, SC23-4865
- *HACMP for AIX 5L: Troubleshooting Guide,* SC23-5177
- *HACMP on Linux: Installation and Administration Guide,* SC23-5211
- *HACMP for AIX 5L: Smart Assist Developer's Guide,* SC23-5210
- *IBM International Program License Agreement.*

HACMP/XD Publications

The HACMP Extended Distance (HACMP/XD) software solutions for disaster recovery, added to the base HACMP software, enable a cluster to operate over extended distances at two sites. HACMP/XD publications include the following:

- *HACMP/XD for Geographic LVM (GLVM): Planning and Administration Guide,* SA23-1338
- *HACMP/XD for HAGEO Technology: Concepts and Facilities Guide, SC23-1922*
- *HACMP/XD for HAGEO Technology: Planning and Administration Guide,* SC23-1886
- *HACMP/XD for Metro Mirror: Planning and Administration Guide,* SC23-4863.

HACMP Smart Assist Publications

The HACMP Smart Assist software helps you quickly add an instance of certain applications to your HACMP configuration so that HACMP can manage their availability. The HACMP Smart Assist publications include the following:

- *HACMP Smart Assist for DB2 User's Guide, SC23-5179*
- *HACMP Smart Assist for Oracle User's Guide, SC23-5178*

- *HACMP Smart Assist for WebSphere User's Guide, SC23-4877*
- *HACMP for AIX 5L: Smart Assist Developer's Guide, SC23-5210*
- *HACMP Smart Assist Release Notes.*

IBM AIX 5L Publications

The following publications offer more information about IBM technology related to or used by HACMP:

- *RS/6000 SP High Availability Infrastructure*, SG24-4838
- *IBM AIX 5L v.5.3 Security Guide,* SC23-4907
- *IBM Reliable Scalable Cluster Technology for AIX 5L and Linux: Group Services Programming Guide and Reference,* SA22-7888
- *IBM Reliable Scalable Cluster Technology for AIX 5L and Linux: Administration Guide,* SA22-7889
- *IBM Reliable Scalable Cluster Technology for AIX 5L: Technical Reference*, SA22-7890
- *IBM Reliable Scalable Cluster Technology for AIX 5L: Messages*, GA22-7891.

Accessing Publications

Use the following Internet URLs to access online libraries of documentation:

AIX 5L, IBM eServer pSeries, and related products:

http://www.ibm.com/servers/aix/library

AIX 5L v.5.3 publications:

http://www.ibm.com/servers/eserver/pseries/library/

WebSphere Application Server publications:

Search the IBM website to access the WebSphere Application Server Library

DB2 Universal Database Enterprise Server Edition publications:

http://www.ibm.com/cgi-bin/db2www/data/db2/udb/winos2unix/support/v8pubs.d2w/en_main#V8PDF

Tivoli Directory Server publications:

http://publib.boulder.ibm.com/tividd/td/IBMDirectoryServer5.1.html

IBM Welcomes Your Comments

You can send any comments via e-mail to hafeedbk@us.ibm.com. Make sure to include the following in your comment or note:

- Title and order number of this book
- Page number or topic related to your comment.

When you send information to IBM, you grant IBM a nonexclusive right to use or distribute the information in any way it believes appropriate without incurring any obligation to you.

Trademarks

The following terms are trademarks of International Business Machines Corporation in the United States or other countries:

- AFS
- AIX
- AIX 5L
- DFS
- @server
- eServer Cluster 1600
- Enterprise Storage Server
- HACMP
- IBM
- NetView
- pSeries
- RS/6000
- Scalable POWERParallel Systems
- Shark
- SP
- xSeries
- WebSphere
- Red Hat Enterprise Linux (RHEL)
- SUSE Linux Enterprise Server
- RPM Package Manager for Linux and other Linux trademarks.

UNIX is a registered trademark in the United States and other countries and is licensed exclusively through The Open Group.

Linux is a registered trademark in the United States and other countries and is licensed exclusively through the GNU General Public License.

Other company, product, and service names may be trademarks or service marks of others.

Chapter 1: Administering an HACMP Cluster

This chapter provides a list of the tasks you perform to configure, maintain, monitor, and troubleshoot an HACMP system, related administrative tasks, and a list of AIX 5L files modified by HACMP.

The main sections in this chapter include:

- Options for Configuring an HACMP Cluster
- Configuration Tasks
- Maintaining an HACMP Cluster
- Monitoring the Cluster
- Troubleshooting an HACMP Cluster
- Related Administrative Tasks
- AIX 5L Files Modified by HACMP
- HACMP Scripts.

Options for Configuring an HACMP Cluster

In HACMP, you can configure a cluster using one of the following HACMP tools:

- HACMP SMIT user interface.
- *WebSMIT utility*. For information on using this utility, see Chapter 2: Administering a Cluster Using WebSMIT.
- *Online Planning Worksheets (OLPW).* This tool provides a convenient method for documenting your cluster configuration: You can use the tool to configure a new cluster or to document an existing cluster. For instructions, see the chapter on Using Online Planning Worksheets in the *Planning Guide*.
- *Two-Node Cluster Configuration Assistant.* Use this tool to configure a basic two-node HACMP cluster. You supply the minimum information required to define a cluster, and HACMP discovers the remainder of the information for you. See the section on Using the Two-Node Cluster Configuration Assistant in the chapter on Creating a Basic HACMP Cluster in the *Installation Guide.*
- *General Configuration Smart Assist.* Start with your installed application and configure a basic cluster (any number of nodes). If you are configuring a WebSphere, DB2 UDB or Oracle application, see the corresponding HACMP Smart Assist guide. See Chapter 3: Configuring an HACMP Cluster (Standard).
- *Cluster Snapshot Utility*. If you have a snapshot of the HACMP cluster configuration taken before an upgrade to HACMP 5.4, you can use the Cluster Snapshot utility to perform the initial configuration. For more information, see Chapter 18: Saving and Restoring Cluster Configurations.

Configuration Tasks

The HACMP configuration tasks are described in detail in subsequent chapters. You can choose to use either the standard or the extended path for the initial configuration, although the standard configuration path is recommended. The major steps in the process are:

- First, configure the cluster topology, and then HACMP resources and resource groups using the standard configuration path

  or

  First, configure the cluster topology, and then HACMP resources and resource groups using the extended configuration path.
- (*Optional*) Configure pre- and post-events, remote notification, HACMP File Collections, cluster verification with automatic corrective action, and other optional settings.
- Verify and synchronize the HACMP configuration.
- Test the cluster.

Configuring HACMP Using the Standard Configuration Path

Using the options under the **Initialization and Standard Configuration** SMIT menu, you can add the basic components of the HACMP cluster to the HACMP Configuration Database (ODM) in a few steps. This configuration path significantly automates the discovery and selection of configuration information and chooses default behaviors.

The prerequisites and default settings of this path are:

- Connectivity for communication must already be established between all cluster nodes. Automatic discovery of cluster information runs by default. That is, once you have configured communication interfaces/devices and established communication paths to other nodes, HACMP automatically collects HACMP-related information and automatically configures the cluster nodes and networks based on physical connectivity. All discovered networks are added to the cluster configuration. This helps you in the configuration process.

  To understand how HACMP maintains the security of incoming connections, see the section Managing Cluster Security and Inter-Node Communications in this chapter.
- IP aliasing is used as the *default* mechanism for binding service IP labels/addresses to network interfaces. For more information, see the chapter on Planning Cluster Network Connectivity in the *Planning Guide*.
- You can configure the most common types of resources. Customization of resource group fallover/fallback behavior supports the most common scenarios.

Chapter 3: Configuring an HACMP Cluster (Standard) takes you through the configuration process if you plan to use the **Initialization and Standard Configuration** path in SMIT. Once you have configured the basic components, you can use the **Extended Configuration** path to customize your configuration.

Configuring HACMP Using the Extended Configuration Path

In order to configure the less common HACMP elements, or if connectivity to each of the cluster nodes is unavailable, you can manually enter the information. When using the menu panels under the **Extended Configuration** SMIT path, if any components are on remote nodes, you must manually initiate the discovery of cluster information. That is, the discovery process used by HACMP is optional when using this path (rather than automatic, as it is when using the **Initialization and Standard Configuration** SMIT path).

Using the options under the **Extended Configuration** SMIT menu, you can add basic components to the HACMP Configuration Database (ODM), as well as additional types of resources. Use the Extended Configuration path to customize the cluster for all the components, policies, and options that are not included in the standard configuration menus.

Configuring Topology and Resources

Chapter 3: Configuring an HACMP Cluster (Standard), describes all the SMIT menus and options available for configuring cluster topology and all the various types of resources supported by the software.

There is an option to configure a distribution preference for the aliases of the service IP labels that are placed under HACMP control. A *distribution preference for service IP label aliases* is a network-wide attribute used to control the placement of the service IP label aliases on the physical network interface cards on the cluster nodes.

For more information, see the section Distribution Preference for Service IP Label Aliases: Overview in Chapter 4: Configuring HACMP Cluster Topology and Resources (Extended).

Configuring Resource Groups and Assigning Resources

Chapter 5: Configuring HACMP Resource Groups (Extended) describes how to configure different types of resource groups. The **Extended Configuration** menus include options for configuring various runtime policies for resource groups as well as for customizing fallover, fallback and startup behavior. It also includes the procedure for adding resources to a resource group.

Configuring Dynamic LPAR and Capacity Upgrade on Demand Resources

Appendix E: Using DLPAR and CUoD in an HACMP Cluster describes how to plan, integrate, configure, and troubleshoot application provisioning for HACMP through the use of dynamic LPAR (DLPAR) and Capacity Upgrade on Demand (CUoD) functions available on some pSeries servers. It also includes examples and recommendations about customizing your existing pre- and post-event scripts.

Configuring Cluster Events

The HACMP system is event-driven. An event is a change of status within a cluster. When the Cluster Manager detects a change in cluster status, it executes the designated script to handle the event and initiates any user-defined customized processing.

To configure customized cluster events, you indicate the script that handles the event and any additional processing that should accompany an event. Chapter 6: Configuring Cluster Events describes the procedures for customization of event handling in HACMP.

Configuring Remote Notification for Cluster Events

The remote notification function allows you to direct SMS text-message notifications to any address including your cell phone.

With previous versions of HACMP, you could alter event scripts to send email when connected to the Internet. Alternately, the remote notification subsystem could send numeric or alphanumeric pages through a dialer modem, which uses the standard Telocator Alphanumeric Protocol (TAP) protocol.

For more information, see the section Defining a New Remote Notification Method in Chapter 6: Configuring Cluster Events.

Verifying and Synchronizing the Configuration

Verifying the cluster configuration assures you that all resources used by HACMP are validly configured, and that ownership and takeover of those resources are defined and are in agreement across all nodes. By default, if the verification is successful, the configuration is automatically synchronized. You should verify the configuration after making changes to a cluster or node. Chapter 7: Verifying and Synchronizing an HACMP Cluster, describes the SMIT menus for verification, explains the contents and uses of the **clverify.log** file, and describes how to verify your cluster.

Chapter 7: Verifying and Synchronizing an HACMP Cluster also explains how to create and maintain HACMP File Collections. Using the HACMP File Collections utility, you can request that a list of files is automatically kept synchronized across the cluster. You no longer have to manually copy an updated file to every cluster node, verify that the file is properly copied, and confirm that each node has the same version of it. If you use the HACMP File Collections utility, HACMP can detect and warn you if one or more files in a collection is deleted or has a zero value on one or more cluster nodes during cluster verifications.

Testing the Cluster

HACMP includes the Cluster Test Tool to help you test the recovery procedures for a new cluster before the cluster becomes part of your production environment. You can also use the tool to test configuration changes in an existing cluster, when the cluster services are not running. Chapter 8: Testing an HACMP Cluster explains how to use the Cluster Test Tool.

Maintaining an HACMP Cluster

The following maintenance tasks for an HACMP system are described in detail in subsequent chapters:

- Starting and Stopping Cluster Services
- Maintaining Shared Logical Volume Manager Components
- Managing the Cluster Topology
- Managing Cluster Resources
- Managing Cluster Resource Groups
- Managing Users and Groups in a Cluster
- Managing Cluster Security and Inter-Node Communications

- Understanding the /usr/es/sbin/cluster/etc/rhosts File
- Saving and Restoring HACMP Cluster Configurations
- Additional HACMP Maintenance Tasks.

Starting and Stopping Cluster Services

Various methods for starting and stopping cluster services are available. Chapter 9: Starting and Stopping Cluster Services describes how to start and stop HACMP on server and client nodes.

Maintaining Shared Logical Volume Manager Components

Any changes to logical volume components must be synchronized across all nodes in the cluster. Chapter 11: Managing Shared LVM Components, and Chapter 12: Managing Shared LVM Components in a Concurrent Access Environment describe how to maintain cluster LVM components. Using C-SPOC (the Cluster Single Point of Control) to configure the cluster components on one node and then synchronize the cluster saves you time and effort.

Managing the Cluster Topology

Any changes to cluster topology require updating the cluster across all nodes. Chapter 13: Managing the Cluster Topology describes how to modify cluster topology after the initial configuration. You can make most changes on one node and then synchronize the cluster.

This chapter also includes information about the **HACMP Communication Interface Management** SMIT menu that lets you configure communication interfaces/devices to AIX 5L without leaving HACMP SMIT.

Managing Cluster Resources

Any changes to cluster resources require updating the cluster across all nodes. You can make most changes on one node and then synchronize the cluster. Chapter 14: Managing the Cluster Resources describes how to modify cluster resources after the initial configuration.

Managing Cluster Resource Groups

Chapter 15: Managing Resource Groups in a Cluster describes how to modify cluster resource groups after the initial configuration. You can add or delete resources and change the runtime policies of resource groups.

You can dynamically migrate resource groups to other nodes and take them online or offline, using the Resource Group Management utility **(clRGmove)** from the command line or through SMIT.

Managing Users and Groups in a Cluster

HACMP lets you manage user accounts for a cluster from a Single Point of Control (C-SPOC). Use C-SPOC to create, change, or remove users and groups from all cluster nodes by executing a C-SPOC command on any single cluster node.

For information, see Chapter 16: Managing User and Groups.

Managing Cluster Security and Inter-Node Communications

You can protect access to your HACMP cluster by setting up security for cluster communications between nodes. HACMP provides security for connections between nodes, with higher levels of security for inter-node communications provided through Kerberos (on SP nodes only) or through virtual private networks (VPN). In addition, you can configure authentication and encryption of the messages sent between nodes.

For information, see Chapter 17: Managing Cluster Security.

Understanding the /usr/es/sbin/cluster/etc/rhosts File

This section explains how and when HACMP uses the **/usr/es/sbin/cluster/etc/rhosts** file, which HACMP uses for inter-node communications. It also describes how this file relates to the **~/.rhosts** file.

The /usr/es/sbin/cluster/etc/rhosts file

A Cluster Communications daemon (**clcomd**) runs on each HACMP node to transparently manage inter-node communications for HACMP. In other words, HACMP manages connections for you automatically:

If the **/usr/es/sbin/cluster/etc/rhosts** file is empty (this is the initial state of this file, upon installation), then **clcomd** accepts the first connection from another node and adds entries to the **/etc/rhosts** file. Since this file is empty upon installation, the first connection from another node adds IP addresses to this file. The first connection usually is performed for verification and synchronization purposes, and this way, for all subsequent connections, HACMP already has entries for node connection addresses in its Configuration Database.

clcomd validates the addresses of the incoming connections to ensure that they are received from a node in the cluster. The rules for validation are based on the presence and contents of the **/usr/es/sbin/cluster/etc/rhosts** file.

- In addition, HACMP includes in the **/usr/es/sbin/cluster/etc/rhosts** file the addresses for all network interface cards from the communicating nodes.
- If the **/usr/es/sbin/cluster/etc/rhosts** file is not empty, then **clcomd** compares the incoming address with the addresses/labels found in the HACMP Configuration Database (ODM) and then in the **/usr/es/sbin/cluster/etc/rhosts** file and allows only listed connections. In other words, after installation, HACMP accepts connections from another HACMP node and adds the incoming address(es) to the local file, thus allowing you to configure the cluster without ever editing the file directly.
- If the **/usr/es/sbin/cluster/etc/rhosts** file is not present, **clcomd** rejects all connections

Typically, you do not manually add entries to the **/usr/es/sbin/cluster/etc/rhosts** file unless you have specific security needs or concerns.

If you are especially concerned about network security (for instance, you are configuring a cluster on an unsecured network), then prior to configuring the cluster, you may wish to manually add all the IP addresses/labels for the nodes to the empty **/usr/es/sbin/cluster/etc/rhosts** file. For information on how to do it, see Manually Configuring /usr/es/sbin/cluster/etc/rhosts file on Individual Nodes in Chapter 17: Managing Cluster Security.

After you synchronize the cluster, you can empty the **/usr/es/sbin/cluster/etc/rhosts** file (but not remove it), because the information present in the HACMP Configuration Database would be sufficient for all future connections.

If the configuration for AIX 5L adapters was changed after the cluster has been synchronized, HACMP may issue an error. See the section Troubleshooting the Cluster Communications Daemon, or the *Troubleshooting Guide* for information on refreshing the **clcomd** utility and updating **/usr/es/sbin/cluster/etc/rhosts**.

The ~/.rhosts File

~/.rhosts is only needed during the migration from pre-5.1 versions of HACMP. Once migration is completed, we recommend removing **~/.rhosts**, if no other applications need **rsh** for inter-node communication.

Saving and Restoring HACMP Cluster Configurations

After you configure the topology and resources of a cluster, you can save the cluster configuration by taking the cluster snapshot. This saved configuration can later be used to restore the configuration if this is needed by applying the cluster snapshot. A cluster snapshot can also be applied to an active cluster to dynamically reconfigure the cluster. Chapter 18: Saving and Restoring Cluster Configurations describes how to use the Cluster Snapshot utility.

Additional HACMP Maintenance Tasks

Additional tasks that you can perform to maintain an HACMP system include changing the log file attributes for a node and performance tuning. For information on these tasks, see the section on Troubleshooting an HACMP Cluster in this chapter.

Monitoring the Cluster

By design, failures of components in the cluster are handled automatically, but you need to be aware of all such events. Chapter 10: Monitoring an HACMP Cluster describes various tools you can use to check the status of an HACMP cluster, the nodes, networks, and resource groups within that cluster, and the daemons that run on the nodes.

The HACMP software includes the Cluster Information Program (Clinfo), based on SNMP. The HACMP for AIX software provides the HACMP for AIX 5L MIB, associated with and maintained by HACMP. Clinfo retrieves this information from the HACMP for AIX Management Information Base (MIB).

The Cluster Manager gathers information relative to cluster state changes of nodes and interfaces. The Cluster Information Program (Clinfo) gets this information from the Cluster Manager and allows clients communicating with Clinfo to be aware of a cluster's state changes. This cluster state information is stored in the HACMP MIB.

Clinfo runs on cluster server nodes and on HACMP client machines. It makes information about the state of an HACMP cluster and its components available to clients and applications via an application programming interface (API). Clinfo and its associated APIs enable you to write applications that recognize and respond to changes within a cluster.

The Clinfo program, the HACMP MIB, and the APIs are described in the *Programming Client Applications Guide*.

Although the combination of HACMP and the high availability features built into the AIX 5L system keeps single points of failure to a minimum, there are still failures that, although detected, can cause other problems.

For suggestions on customizing error notification for various problems not handled by the HACMP events, the *Planning Guide*.

Troubleshooting an HACMP Cluster

It is useful to follow guidelines for troubleshooting. You should be aware of all the diagnostic tools available from HACMP and AIX 5L. See Chapter 1: Troubleshooting HACMP Clusters in the *Troubleshooting Guide* for suggested troubleshooting guidelines, as well as for information on tuning the cluster for best performance.

When you become aware of a problem, the first place to look for helpful diagnostic information is the log files. Chapter 2: Using Cluster Log Files in the *Troubleshooting Guide* describes how to use the various log files. This chapter also contains information on viewing and maintaining log file parameters and instructions for redirecting log files.

If log files do not help you resolve the issue, you may need to check cluster components. See Chapter 3: Investigating System Components and Solving Common Problems in the *Troubleshooting Guide* for suggested strategies as well as for a list of solutions to common problems that may occur in an HACMP environment.

For information specific to Reliable Scalable Cluster Technology (RSCT) subsystems, see the following IBM publications:

- *IBM Reliable Scalable Cluster Technology for AIX 5L and Linux: Group Services Programming Guide and Reference,* SA22-7888
- *IBM Reliable Scalable Cluster Technology for AIX 5L and Linux: Administration Guide,* SA22-7889
- *IBM Reliable Scalable Cluster Technology for AIX 5L: Technical Reference*, SA22-7890
- *IBM Reliable Scalable Cluster Technology for AIX 5L: Messages*, GA22-7891

Related Administrative Tasks

The tasks below, while not specifically discussed in this book, are essential for effective system administration.

Backing Up Your System

The practice of allocating multiple copies of a logical volume can enhance high availability in a cluster environment, but it should not be considered a replacement for regular system backups. Although HACMP is designed to survive failures within the cluster, it cannot survive a catastrophic failure where multiple points of failure leave data on disks unavailable. Therefore, to ensure data reliability and to protect against catastrophic physical volume failure, you must have a backup procedure in place and perform backups of your system on a regular basis.

To maintain your HACMP environment, you must back up the root volume group (which contains the HACMP software) and the shared volume groups (which contain the data for highly available applications) regularly. HACMP is like other AIX 5L environments from this perspective. Back up all nodes.

Documenting Your System

As your HACMP system grows and changes, it differs from its initial cluster configuration. It is your responsibility as system administrator to document all aspects of the HACMP system unique to your environment. This responsibility includes documenting procedures concerning the highly available applications, recording changes that you make to the configuration scripts distributed with HACMP, documenting any custom scripts you write, recording the status of backups, maintaining a log of user problems, and maintaining records of all hardware. This documentation, along with the output of various display commands and cluster snapshots, will be useful for you, as well as for IBM support, to help resolve problems.

Starting with HACMP 5.2, you can use the report supplied with the Online Planning Worksheet program to generate a of a cluster configuration, then save and print the report to document the system.

Maintaining Highly Available Applications

As system administrator, you should understand the relationship between your applications and HACMP. To keep the applications highly available, HACMP starts and stops the applications that are placed under HACMP control in response to cluster events. Understanding when, how, and why this happens is critical to keeping the applications highly available, as problems can occur that require corrective actions.

For a discussion of strategies for making your applications highly available, see the planning chapters and Appendix B on Applications and HACMP in the *Planning Guide*.

Helping Users

As the resident HACMP expert, you can expect to receive many questions from end users at your site about HACMP. The more you know about HACMP, the better you are able to answer these questions. If you cannot answer questions about your HACMP cluster environment, contact your IBM support representative.

AIX 5L Files Modified by HACMP

The following AIX 5L files are modified to support HACMP. They are not distributed with HACMP.

/etc/hosts

The cluster event scripts use the **/etc/hosts** file for name resolution. All cluster node IP interfaces must be added to this file on each node.

HACMP may modify this file to ensure that all nodes have the necessary information in their **/etc/hosts** file, for proper HACMP operations.

If you delete service IP labels from the cluster configuration using SMIT, we recommend that you also remove them from **/etc/hosts**. This reduces the possibility of having conflicting entries if the labels are reused with different addresses in a future configuration.

Note that DNS and NIS are disabled during HACMP-related name resolution. This is why HACMP IP addresses must be maintained locally.

/etc/inittab

The **/etc/inittab** file is modified in each of the following cases:

- HACMP is configured for IP address takeover
- The **Start at System Restart** option is chosen on the SMIT **System Management (C-SPOC) > Manage HACMP Services > Start Cluster Services** panel
- Concurrent Logical Volume Manager (CLVM) is installed with HACMP
- Starting with HACMP 5.3, the **/etc/inittab** file has the following entry in the **/user/es/sbin/cluster/etc/rc.init**:

  ```
  hacmp:2:once:/usr/es/sbin/cluster/etc/rc.init
  ```

  This entry starts the HACMP Communications Daemon, **clcomd**, and the **clstrmgr** subsystem.

Modifications to the /etc/inittab File due to IP Address Takeover

The following entry is added to the **/etc/inittab** file for HACMP network startup with IP address takeover:

```
harc:2:wait:/usr/es/sbin/cluster/etc/harc.net # HACMP network startup
```

When IP address takeover is enabled, the system edits **/etc/inittab** to change the **rc.tcpip** and **inet**-dependent entries from run level “2” (the default multi-user level) to run level “a”. Entries that have run level “a” are processed only when the **telinit** command is executed specifying that specific run level.

Modifications to the /etc/inittab File due to System Boot

The **/etc/inittab** file is used by the **init** process to control the startup of processes at boot time.

When the system boots, the **/etc/inittab** file calls the **/usr/es/sbin/cluster/etc/rc.cluster** script to start HACMP. The entry is added to the **/etc/inittab** file if the **Start at system restart** option is chosen on the SMIT **System Management (C-SPOC) > Manage HACMP Services > Start Cluster Services** panel or when the system boots:

```
hacmp:2:once:/usr/es/sbin/cluster/etc/rc.init
```

This starts the HACMP Communications Daemon, **clcomd**, and the **clstrmgr** subsystem.

Because the **inet** daemons must not be started until after HACMP-controlled interfaces have swapped to their service IP address, HACMP also adds the following entry to the end of the **/etc/inittab** file to indicate that **/etc/inittab** processing has completed:

```
clinit:a:wait:/bin/touch /usr/es/sbin/cluster/.telinit
#HACMP for AIX These must be the last entry in run level "a" in inittab!
pst_clinit:a:wait:/bin/echo Created /usr/es/sbin/cluster/ .telinit >
/dev/console
#HACMP for AIX These must be the last entry in run level "a" in inittab!
```

See Chapter 9: Starting and Stopping Cluster Services, for more information about the files involved in starting and stopping HACMP.

/etc/rc.net

The **/etc/rc.net** file is called by **cfgmgr**, (**cfgmgr** is the AIX 5L utility that configures devices and optionally installs device software into the system), to configure and start TCP/IP during the boot process. It sets hostname, default gateway, and static routes. The following entry is added at the beginning of the file for a node on which IP address takeover is enabled:

```
# HACMP for AIX \
# HACMP for AIX These lines added by HACMP for AIX software
[ "$1" = "-boot" ] && shift || {              # HACMP for AIX
    ifconfig lo0 127.0.0.1 up;                # HACMP for AIX
    /bin/uname -S`hostname|sed 's/\..*$//'`; # HACMP for AIX
    exit 0;                                   # HACMP for AIX
}                                             # HACMP for AIX
#
```

The HACMP entry prevents **cfgmgr** from reconfiguring boot and service IP addresses while HACMP is running.

/etc/services

The **/etc/services** file defines the sockets and protocols used for network services on a system. The ports and protocols used by the HACMP components are defined here.

```
#clinfo_deadman                 6176/tcp
#clsmuxpd                       6270/tcp
#clm_lkm                        6150/tcp
#clm_smux                       6175/tcp
```

```
#godm                           6177/tcp
#topsvcs                        6178/udp
#grpsvcs                        6179/udp
#emsvcs                         6180/udp
#clver                          6190/tcp
#clcomd                         6191/tcp
```

Note: If, in addition to HACMP, you install HACMP/XD for GLVM, the following entry for the port number and connection protocol is automatically added to the **/etc/services** file on each node on the local and remote sites on which you installed the software: `rpv 6192/tcp.` This default value enables the RPV server and RPV client to start immediately after they are configured, that is, to be in the *available state*. For more information, see *HACMP/XD for GLVM Planning and Administration Guide.*

/etc/snmpd.conf

Note: The default version of the **snmpd.conf** file for AIX 5L v.5.2 and v. 5.3 is **snmpdv3.conf**.

The SNMP daemon reads the **/etc/snmpd.conf** configuration file when it starts up and when a **refresh** or **kill -1** signal is issued. This file specifies the community names and associated access privileges and views, hosts for trap notification, logging attributes, **snmpd**-specific parameter configurations, and SMUX configurations for the **snmpd**. The HACMP installation process adds a **clsmuxpd** password to this file. The following entry is added to the end of the file, to include the HACMP MIB supervised by the Cluster Manager:

```
smux  1.3.6.1.4.1.2.3.1.2.1.5 "clsmuxpd_password" # HACMP clsmuxpd
```

HACMP supports SNMP Community Names other than “public.” That is, HACMP will function correctly if the default SNMP Community Name has been changed in **/etc/snmpd.conf** to be anything other than “public” (the default). The SNMP Community Name used by HACMP is the first name found that is not “private” or “system” using the **lssrc -ls snmpd** command.

The Clinfo service also gets the SNMP Community Name in the same manner. The Clinfo service supports the **-c** option for specifying SNMP Community Name but its use is not required. The use of the **-c** option is considered a security risk because doing a **ps** command could find the SNMP Community Name. If it is important to keep the SNMP Community Name protected, change permissions on **/tmp/hacmp.out**, **/etc/snmpd.conf**, **/smit.log** and **/usr/tmp/snmpd.log** to not be world readable.

Note: See the AIX documentation for full information on the **/etc/snmpd.conf** file. Version 3 (default for AIX 5.2 and up) has some differences from Version 1.

/etc/snmpd.peers

The **/etc/snmpd.peers** file configures **snmpd** SMUX peers. During installation, HACMP adds the following entry to include the **clsmuxpd** password to this file:

```
clsmuxpd 1.3.6.1.4.1.2.3.1.2.1.5 "clsmuxpd_password" # HACMP clsmuxpd
```

/etc/syslog.conf

The **/etc/syslog.conf** configuration file is used to control output of the **syslogd** daemon, which logs system messages. During the install process HACMP adds entries to this file that direct the output from HACMP-related problems to certain files.

```
# example:
# "mail messages, at debug or higher, go to Log file. File must exist."
# "all facilities, at debug and higher, go to console"
# "all facilities, at crit or higher, go to all users"
#  mail.debug           /usr/spool/mqueue/syslog
#  *.debug              /dev/console
#  *.crit                       *
# HACMP Critical Messages from HACMP
local0.crit /dev/console
# HACMP Informational Messages from HACMP
local0.info /usr/es/adm/cluster.log
# HACMP Messages from Cluster Scripts
user.notice /usr/es/adm/cluster.log
# HACMP/ES for AIX Messages from Cluster Daemons
daemon.notice /usr/es/adm/cluster.log
```

The **/etc/syslog.conf** file should be identical on all cluster nodes.

/etc/trcfmt

The **/etc/trcfmt** file is the template file for the system trace logging and report utility, **trcrpt**. The installation process adds HACMP tracing to the trace format file. HACMP tracing is performed for the **clstrmgr** and **clinfo** daemons.

Note: HACMP 5.3 and up no longer uses the **clsmuxpd** daemon; the SNMP server functions are included in the Cluster Manager—the **clstrmgr** daemon.

/var/spool/cron/crontab/root

The **/var/spool/cron/crontab/root** file contains commands needed for basic system control. The installation process adds HACMP logfile rotation to the file.

HACMP Scripts

The HACMP software contains the following scripts.

Startup and Shutdown Scripts

The HACMP software uses each of the following scripts during starting and stopping the cluster services:

/usr/es/sbin/cluster/utilities/clstart

The **/usr/es/sbin/cluster/utilities/clstart** script, which is called by the **/usr/es/sbin/cluster/etc/rc.cluster** script, invokes the AIX 5L System Resource Controller (SRC) facility to start the cluster daemons. The **clstart** script starts HACMP with the options currently specified on the **System Management (C-SPOC) > Manage HACMP Services > Start Cluster Services** SMIT panel.

There is a corresponding C-SPOC version of this script that starts cluster services on each cluster node. The **/usr/es/sbin/cluster/sbin/cl_clstart** script calls the HACMP **clstart** script.

At cluster startup, **clstart** looks for the file **/etc/rc.shutdown**. The system file **/etc/rc.shutdown** can be configured to run user-specified commands during processing of the AIX 5L **/usr/sbin/shutdown** command.

Newer versions of the AIX 5L **/usr/sbin/shutdown** command automatically call HACMP's **/usr/es/sbin/cluster/etc/rc.shutdown**, and subsequently call the existing **/etc/rc.shutdown** (if it exists).

Older versions of the AIX 5L **/usr/sbin/shutdown** command do not have this capability. In this case, HACMP manipulates the /**etc/rc.shutdown** script, so that both **/usr/es/sbin/cluster/etc/rc.shutdown** and the existing **/etc/rc.shutdown** (if it exists) are run. Since HACMP needs to stop cluster services before the **shutdown** command is run, on cluster startup, **rc.cluster** replaces any user supplied **/etc/rc.shutdown** file with the HACMP version. The user version is saved and is called by the HACMP version prior to its own processing. When cluster services are stopped, the **clstop** command restores the user's version of **rc.shutdown**.

/usr/es/sbin/cluster/utilities/clstop

The **/usr/es/sbin/cluster/utilities/clstop** script, which is called from the SMIT **Stop Cluster Services** panel, invokes the SRC facility to stop the cluster daemons with the options specified on the **Stop Cluster Services** panel.

There is a corresponding C-SPOC version of this script that stops cluster services on each cluster node. The **/usr/es/sbin/cluster/sbin/cl_clstop** script calls the HACMP **clstop** script.

Also see the notes on **/etc/rc.shutdown** in the section on **clstart** above for more information.

/usr/es/sbin/cluster/utilities/clexit.rc

If the SRC detects that the **clstrmgr** daemon has exited abnormally, it executes the **/usr/es/sbin/cluster/utilities/clexit.rc** script to halt the system. If the SRC detects that any other HACMP daemon has exited abnormally, it executes the **clexit.rc** script to stop these processes, but does not halt the node.

You can change the default behavior of the **clexit.rc** script by configuring the **/usr/es/sbin/cluster/etc/hacmp.term** file to be called when the HACMP cluster services terminate abnormally. You can customize the **hacmp.term** file so that HACMP will take actions specific to your installation. See the **hacmp.term** file for full information.

/usr/es/sbin/cluster/etc/rc.cluster

If the **Start at system restart** option is chosen on the **System Management (C-SPOC) > Manage HACMP Services > Start Cluster Services** SMIT panel, the **/usr/es/sbin/cluster/etc/rc.cluster** script is called by the **/etc/inittab** file to start HACMP. The **/usr/es/sbin/cluster/etc/rc.cluster** script does some necessary initialization and then calls the **usr/es/sbin/cluster/utilities/clstart** script to start HACMP.

The **/usr/es/sbin/cluster/etc/rc.cluster** script is also used to start the **clinfo** daemon on a client.

A corresponding C-SPOC version of this script starts cluster services on each cluster node. The **/usr/es/sbin/cluster/sbin/cl_rc.cluster** script calls the HACMP **rc.cluster** script.

See the man page for **rc.cluster** for more information.

/etc/rc.net

The **/etc/rc.net** script is called by the **/usr/es/sbin/cluster/etc/rc.cluster** script to configure and start the TCP/IP interfaces and to set the required network options. The **/etc/rc.net** script is used in the boot process to retrieve interface information from the ODM and to configure all defined interfaces. If IP address takeover is configured, the **/etc/rc.net** script is called from the **/usr/es/sbin/cluster/etc/rc.cluster** script at cluster startup instead of during the boot process.

Event Scripts

The node, network, resource group, server, site, and other event scripts are called by the cluster daemons to respond to cluster events. The event scripts are found in the **/usr/es/sbin/cluster/events** directory.

For more information about these scripts, see the chapter on planning cluster events in the *Planning Guide*, and Chapter 6: Configuring Cluster Events in this Guide.

/usr/es/sbin/cluster/etc/clinfo.rc Script

The **/usr/es/sbin/cluster/etc/clinfo.rc** script, which is invoked by the **clinfo** daemon whenever a network or node event occurs, updates the system's ARP cache. You can customize this script for additional processing. There must be a copy of the **/usr/es/sbin/cluster/etc/clinfo.rc** script on each node in the cluster. See the **clinfo.rc** man page for additional information.

Chapter 2: Administering a Cluster Using WebSMIT

HACMP includes a Web-enabled user interface (WebSMIT) that provides consolidated access to:

- HACMP SMIT configuration and management functions
- Interactive cluster status display
- HACMP online documentation
- Graphical displays of cluster topology and resource group dependencies.
- User Authentication settings (optional)—administrators can specify a group of users that have read-only access. Those users have permissions to view configuration and status, and navigate through SMIT panels, but cannot execute commands or make changes.
- Support for Mozilla-based browsers (Mozilla 1.7.3 for AIX and FireFox 1.0.6) in addition to Internet Explorer versions 6.0 and higher.

The WebSMIT interface includes an interface similar to the ASCII SMIT interface in addition to other information. Because WebSMIT runs in a Web browser, you can access it from any platform.

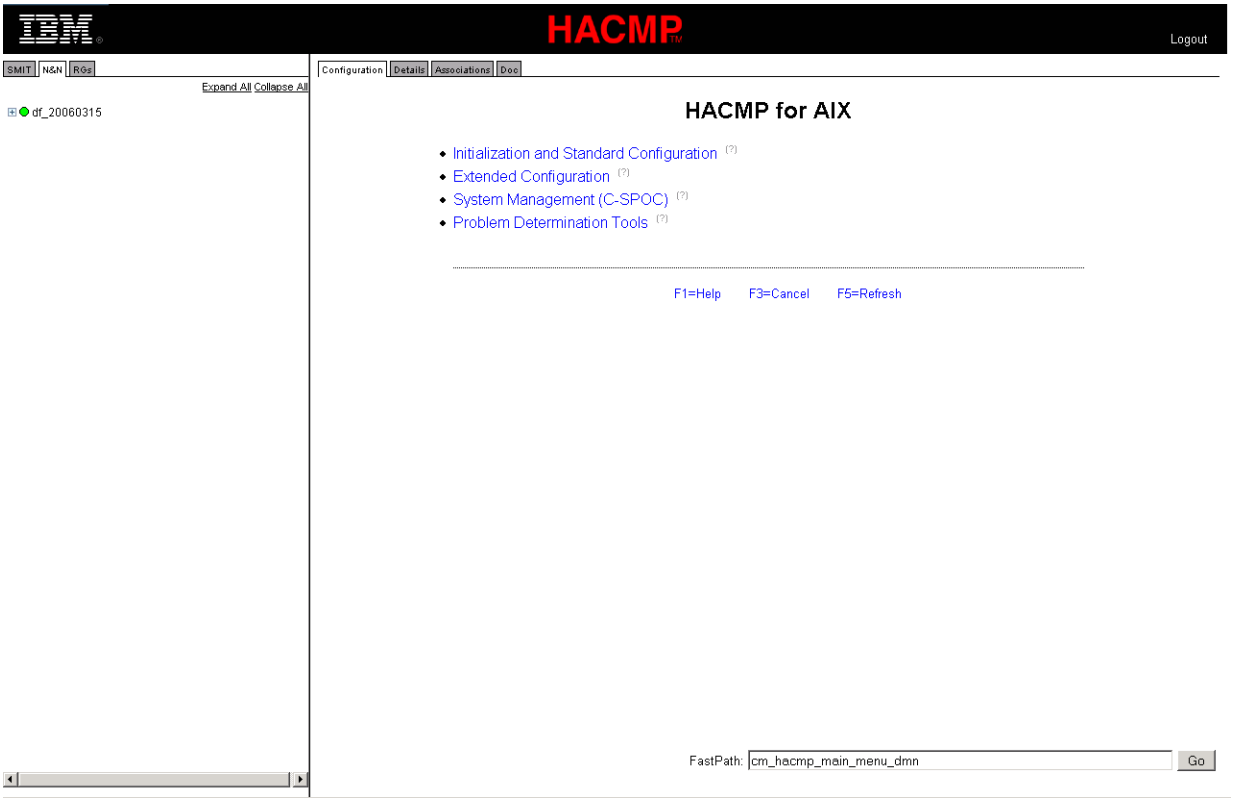
To use the WebSMIT interface, you must configure and run a Web server process on the cluster node(s) to be administered. See the **/usr/es/sbin/cluster/wsm/README** file for information on setting up WebSMIT to work with your web server, the default security mechanisms in place when installing HACMP 5.2 and up, and the configuration files available for customization. For more information about configuring WebSMIT, see the section Installing and Configuring WebSMIT in the *Installation Guide*.

The following sections describe WebSMIT's functionality:

- Working with WebSMIT
- Configuring HACMP Using WebSMIT
- Viewing the Cluster Components
- Viewing Cluster Configuration Information in WebSMIT
- Viewing HACMP Documentation in WebSMIT.

Working with WebSMIT

Here is the WebSMIT main display page:



WebSMIT Display

.

The WebSMIT display is divided into the following frames, which are described in detail in the following sections:

- Header Frame
- Navigation Frame
- Activity Frame

Header Frame

The top frame displays the name of the cluster and provides a logout link.

Navigation Frame

The left frame displays three tabbed views from which you can navigate your cluster, as well as configuration menus. These navigation tabs display items in an expandable, hierarchical view. Selecting an item updates the content displayed in the Activity frame to reflect the current selection. Open and contract the tree as needed to show or hide the sub-components by

clicking on the + or - symbol or by using the **Expand/Collapse All** button. In the previous figure, the WebSMIT Display shows the Navigation frame with the Nodes and Networks tab selected.

You may select the following tabs from the Navigation frame:

- **SMIT** tab. Provides hierarchical navigation of the SMIT menus to configure and manage your cluster. Clicking on a menu item from the **SMIT** tab, displays the corresponding SMIT panel in the Activity frame **on the Configuration** tab.
- **N&N** (Nodes and Networks) tab. Contains an expandable hierarchical view based on the cluster topology (either site- or node-centric depending on the cluster definition).

  The icons to the left of the hierarchical menu items indicates the state of the site, node, or network.

  Clicking on the **N&N** tab updates the tabs available in the **Activity** frame.
- **RGs** (Resource Groups) tab. Provides an expandable hierarchical view, based on the cluster resources. The icons to the left of the hierarchical menu items indicate the state of the corresponding cluster resource.

If a branch of a cluster hierarchy has a flashing status indicator, one of its nested objects has a problem. For example, if a node is in the error state, the site that contains it has a flashing status indicator, and the cluster that contains it likewise has a flashing indicator. By expanding the flashing branches of the cluster hierarchy, you can find where the problem is.

Activity Frame

The pane on the right side displays information corresponding to the items selected in the Navigation frame. You may select the following tabs from the Activity frame:

- **Configuration** tab. Displays items selected in the **SMIT** tab of the Navigation frame.
- **Details** tab. Displays cluster components selected in the **N&N** tab or the **RGs** tab. To view updated configuration information, click the **Refresh** button, located on the upper left of the page, after making changes to the cluster.
  - **Associations** tab. View a graphical display of objects related to the item selected from the **N&N** tab or the **RGs** tab in the Navigation frame.
  - **Docs** tab. View the HACMP for AIX 5L documentation bookshelf page for access to the HACMP documentation installed on your system. This tab also provides a link to the online HACMP for AIX 5L documentation.

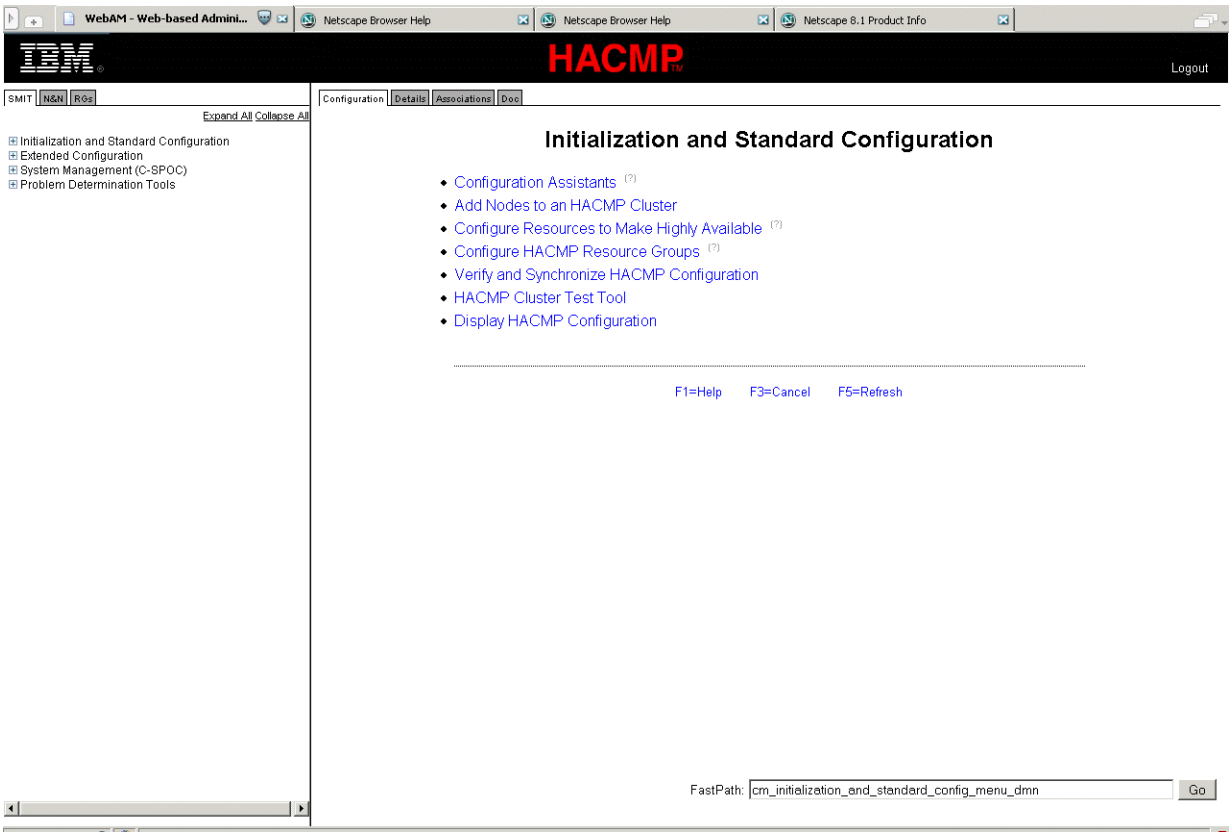
Tooltip Help is available where indicated by a light gray question mark, slightly smaller than the surrounding text: move your mouse over the tooltip indicator to view the help text about possible actions you may take.

Holding your mouse over the help status indicator changes the cursor to the help cursor. To display help on the possible action(s) you may take, right-click on the item

Configuring HACMP Using WebSMIT

Clicking on the **SMIT** tab from the Navigation frame displays the expandable SMIT menus used to configure and manage your cluster. Clicking on a SMIT menu item, displays the corresponding SMIT panel in the Activity frame on the **Configuration** tab.

A fastpath text box is located at the bottom of the display area. You may enter another SMIT panel fastpath in this text box to display that SMIT panel.



WebSMIT: SMIT Tab and HACMP Configuration Tab

Common WebSMIT Panel Options

The common WebSMIT panel options are mapped as follows:

Key or Command	Action
F1 help	Displays context help for the current screen. For menus, a popup menu (Tell me more) displays the help text in a separate window.
F3 cancel	This is mapped to the browser **Back** function.
F4 list	Pressing F4 or selecting the **List** button next to an item creates a popup selection list.
F5 reset	This is a built-in function of the browser.
F6 show command	Displays the command that was created from the input provided.
F7 edit	This function key is *not* mapped.

Key or Command	Action
F8 image	This function key is *not* mapped. There are several built-in functions of the browser that provide these functions.
F9 shell	This function key is *not* mapped.
F10 exit	This function key is *not* mapped.
Enter = do	Runs the command by pressing ENTER.
/ = Find n = Find Next>	These keys are *not* mapped. The browser has built-in search capability.
Fast Paths>	At the bottom of each panel is a text entry where you can enter an HACMP SMIT fast path. The current panel ID is displayed.

Browser Controls

Pressing the browser **stop** button stops the current page from loading; it does *not* stop any commands that are being run on the server.

Pressing the browser **reload** button reloads the current page.

Functional Limitations

SMIT panels that use interactive input, such as entering a password or "Mount volume 2 on cd0 and press ENTER to continue," are *not* supported. WebSMIT displays a default page when you attempt to access these pages directly.

WebSMIT Logs

All operations of the WebSMIT interface are logged to the **wsm_smit.log** file and are equivalent to the logging done with **smitty -v**. Script commands are also captured in the **wsm_smit.script** log file. All logs go to **/usr/es/sbin/cluster/wsm/logs**. The WebSMIT logs are *not* subject to manipulation (redirect, backup) by HACMP logs. Just like **smit.log** and **smit.script**, the files grow indefinitely.

The **snap -e** utility captures the WebSMIT log files only in the default location (**/usr/es/sbin/cluster/wsm/logs**).

Configuring and Managing Nodes and Networks in WebSMIT

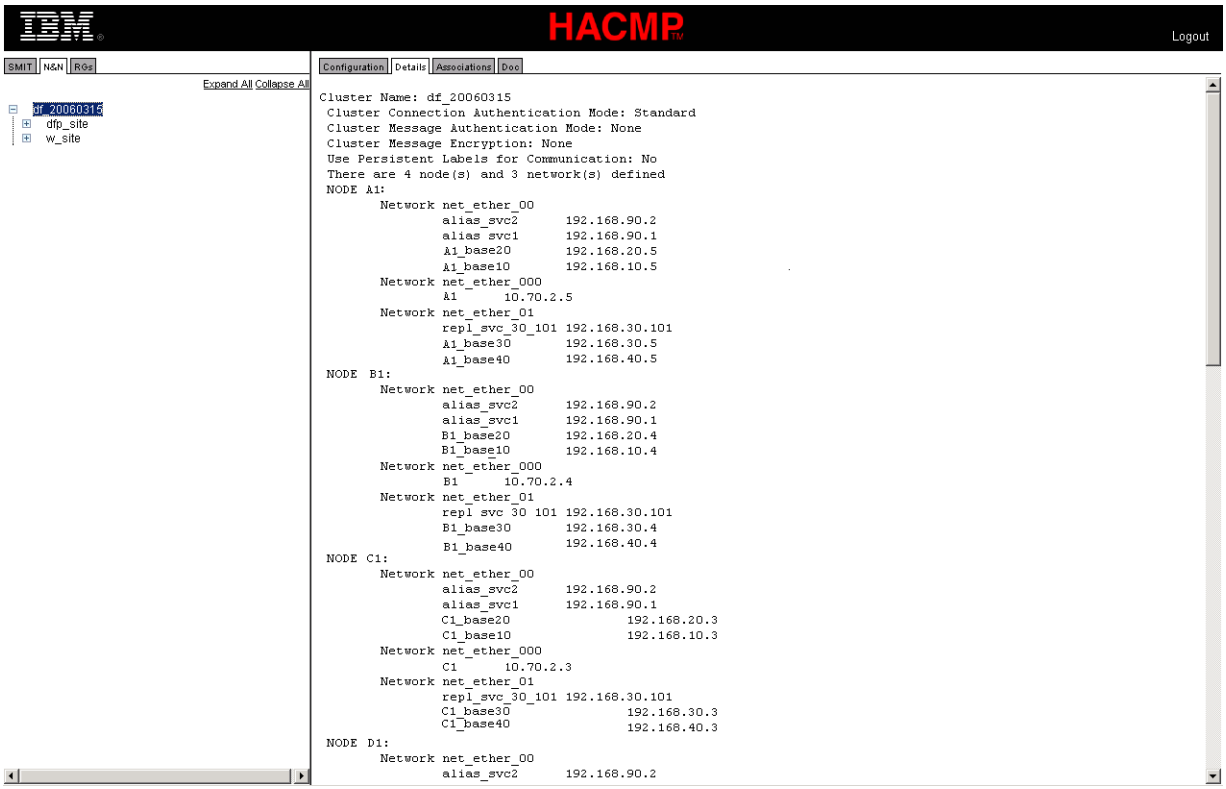
The **N&N** (Nodes and Networks) tab contains an expandable hierarchical view based on the cluster topology, and the status of the cluster sites, nodes, and resource groups. The cluster object displays first, followed by the next item in the hierarchy: **Sites** if one or more sites exist, or **Nodes** otherwise. WebSMIT displays the **N&N** tab by default upon login and updates the contents automatically.

Note: The tree is *not* a representation of "configuration" but of state. Therefore, if the cluster is down, the hierarchy will *not* be shown

To view configuration information of an item selected in the **N&N** tab, select the **Details** tab from the Activity frame on the right.

To view a site or node-centric graphical display of the cluster components, select the **Associations** tab from the Activity frame.

The following figure shows the **N&N** tab displayed in the Navigation frame and the Details Tab displayed in the Activity frame.



WebSMIT: Nodes and Networks View tab and Details tab

WebSMIT Nodes and Networks Status Indicators

Status icons adjacent to the items in the hierarchy on the **N&N** tab indicate the status of each item, according to the icon's color and shape. A flashing icon indicates that a sub-item is in an ERROR state.

Note: You must have cluster services running on at least one node for the hierarchy to display.

The table below shows the status indicator description for each **N&N** tab object.

Object	Circle - Green	Square - Red	Caution Sign- Yellow	Diamond- Blue
Cluster	All cluster resources are up.	• All cluster resources are down. • Cluster Manager has stopped.	Error	N/A
Site	All nodes in the site are up.	All nodes in the site are down.	Error	N/A
Nodes	All nodes in the cluster or site |are up.	All nodes in the cluster or site are down.	Error	All composite's children do *not* have the same state.
Node	All networks and resource groups are up.	All networks and resource groups are down.	Error	Node is in an intermediate state, like JOINING or LEAVING.
Network	All interfaces are up.	All interfaces are down.	Error	N/A
Resource Group online on home node only or first available	Resource group is online and on home node.	Resource group is down.	Error if the resource group is in the ERROR state	Resource group is in an intermediate state, such as, ACQUIRING.
Resource Group online on all available nodes	Resource group is online on all nodes in the resource group.	Resource group is offline on all nodes in the resource group.	Resource group is in the ERROR state on all nodes.	

WebSMIT N&N Tab Right-Click SMIT Options

Right-click on a menu item to display a list of SMIT entries. Selecting a SMIT menu displays the appropriate SMIT panel in the Activity frame on the **Configuration** tab as shown below:

Menu Item	SMIT Entries

Cluster	`Add/Change/Show an HACMP Cluster`
	`Remove an HACMP Cluster`
	`Add a Site`
	`Add Nodes to an HACMP Cluster`
	`Manage HACMP Services`
	`Discover HACMP-related Information from Configured Nodes`
	`Verify and Synchronize HACMP Configuration`

Site	`Add a Site`
	`Change/Show a Site`
	`Remove a Site`

Nodes	`Add a Node to the HACMP Cluster` `Remove a Node in the HACMP Cluster`

Node	`Add a Node to the HACMP Cluster`
	`Change/Show a Node in the HACMP Cluster`
	`Remove a Node in the HACMP Cluster`

Network	`Add a Network to the HACMP Cluster`
	`Change/Show a Network in the HACMP Cluster`
	`Remove a Network from the HACMP Cluster`
	`Add Communication Interfaces/Devices`

Resource Group	`Add a Resource Group`
	`Change/Show a Resource Group`
	`Change/Show Resources and Attributes for a Resource Group`
	`Remove a Resource Group`
	`Bring on/off-line`

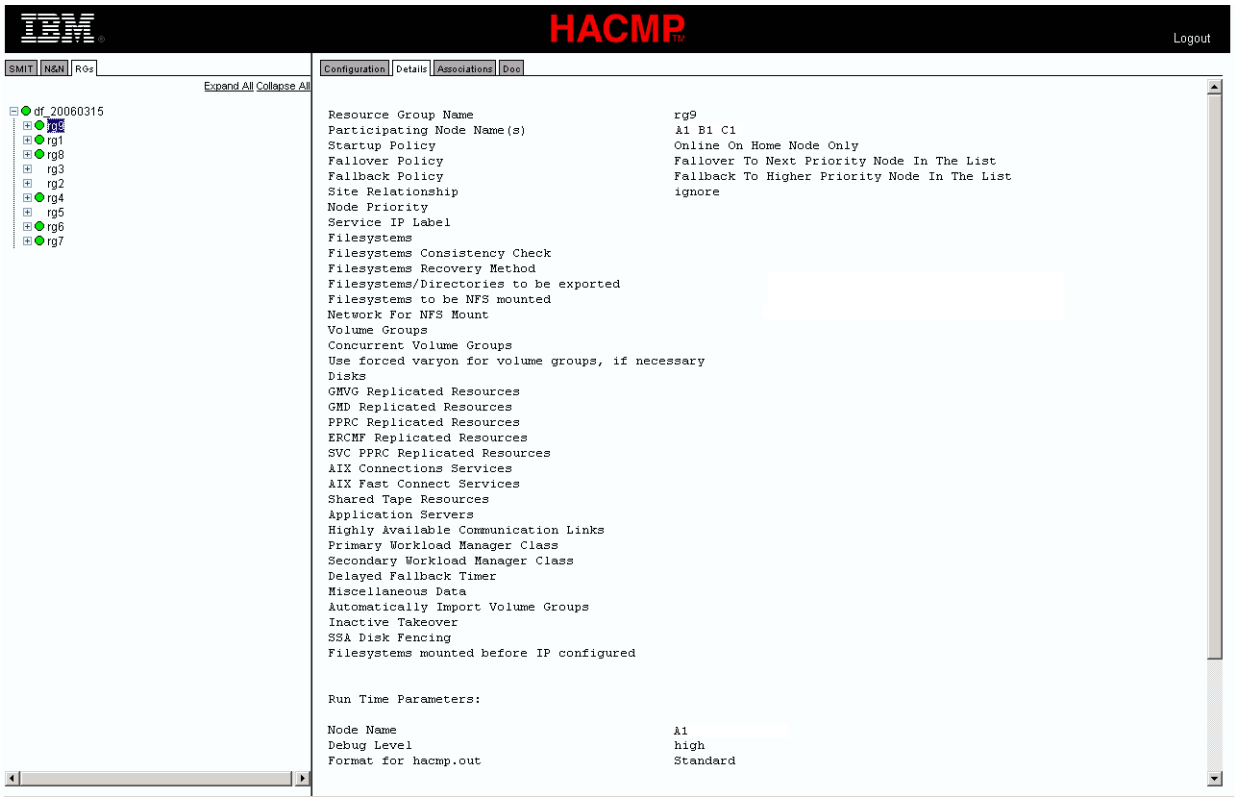
Configuring and Managing Resources in WebSMIT

The **RGs** (Resource Groups) tab contains an expandable hierarchical menu based on the cluster resource groups, and the status of the cluster resource components. The content of this tab updates automatically as changes occur in the cluster.

Selecting a menu item from the **RGs** tab displays its configuration information under the **Details** tab in the Activity frame.

To view a resource group-centric graphical display of the cluster components hierarchy, select the **Associations** tab from the right pane.

The following figure shows the **RGs** tab displayed in the Navigation frame and the Details Tab displayed in the Activity frame.



WebSMIT: RGs View tab and Details tab

WebSMIT RGs Tab Status Indicators

The status icons displayed on the **RGs View** tab indicate the state of the top-level cluster object. The status icons also display adjacent to the items in the tree to indicate the status of the each item. These status indicators give you the state of the cluster object.

Note: You must have cluster services running on at least one node for status icons to display.

The color and shape icon to the left of the text of each item indicates its status. The following table shows the status indicator description for each **RGs** tab objects.

Object	Circle - Green	Square - Red	Caution Sign- Yellow	Diamond- Blue
Cluster	All cluster are resources up.	• All cluster resources are down. or • Cluster Manager is stopped.	Error	N/A
Node Location	Resource group is ONLINE on node.	Resource group is in ERROR state on node.	Error	N/A
Non-concurrent Resource Group	ONLINE on a participating node.	Resource group is DOWN.	Resource group is in the ERROR state.	Resource group is in an intermediate state, like ACQUIRING.
Concurrent Resource Group	ONLINE on all nodes in the resource group.	DOWN state on all participating nodes.	ERROR on all participating nodes	Resource group is not in the same state on all participating nodes (for example, UP on some, DOWN or ERROR on others).
Service IP Label	Service IP label is up on boot interface.	Service IP label is down.	N/A	N/A
Application Server	Started	Stopped	N/A	N/A

Note: In WebSMIT, you cannot run the cluster verification process in an interactive mode.

WebSMIT RGs Tab Right-Click SMIT Options

Right-click on a menu item to display a list of SMIT entries that, when selected, will display the appropriate SMIT panel in the Activity frame on the **Configuration** tab as shown below:

Menu Item	SMIT Entries

Cluster	`Add/Change/Show an HACMP Cluster`
	`Remove an HACMP Cluster`
	`Add a Site`
	`Add Nodes to an HACMP Cluster`
	`Manage HACMP Services`
	`Discover HACMP-related Information from Configured Nodes`
	`Verify and Synchronize HACMP Configuration`

Site	`Add a Site`
	`Change/Show a Site`
	`Remove a Site`

Nodes	`Add a Node to the HACMP Cluster` `Remove a Node in the HACMP Cluster`

Node	`Add a Node to the HACMP Cluster`
	`Change/Show a Node in the HACMP Cluster`
	`Remove a Node in the HACMP Cluster`

Network	`Add a Network to the HACMP Cluster`
	`Change/Show a Network in the HACMP Cluster`
	`Remove a Network from the HACMP Cluster`
	`Add Communication Interfaces/Devices`

Resource Group	`Add a Resource Group`
	`Change/Show a Resource Group`
	`Change/Show Resources and Attributes for a Resource Group`
	`Remove a Resource Group`
	`Bring on/off-line`

Note: In WebSMIT, you cannot run the cluster verification process in an interactive mode.

Viewing the Cluster Components

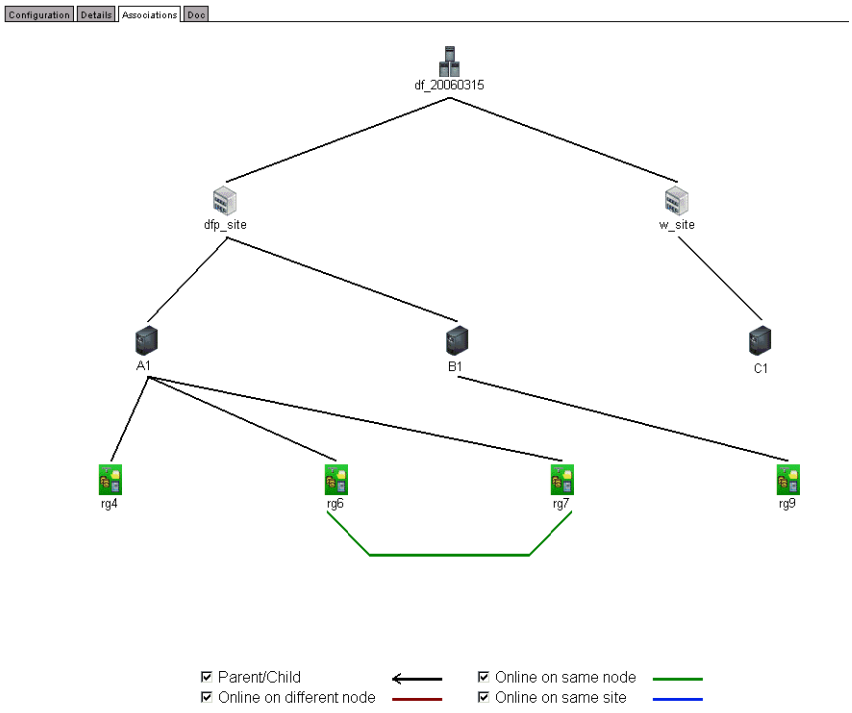
From the Activity frame, click the **Associations** tab to view a graphical display of the cluster components hierarchy corresponding to the item selected in the Navigation frame: the **N&N** tab or the **RGs** tab**.** WebSMIT continuously updates the Navigation frame status icons to indicate state of the current cluster components.

Both the **N&N** view and the **RGs** view show the cluster components in a hierarchical manner. The difference between the two views is:

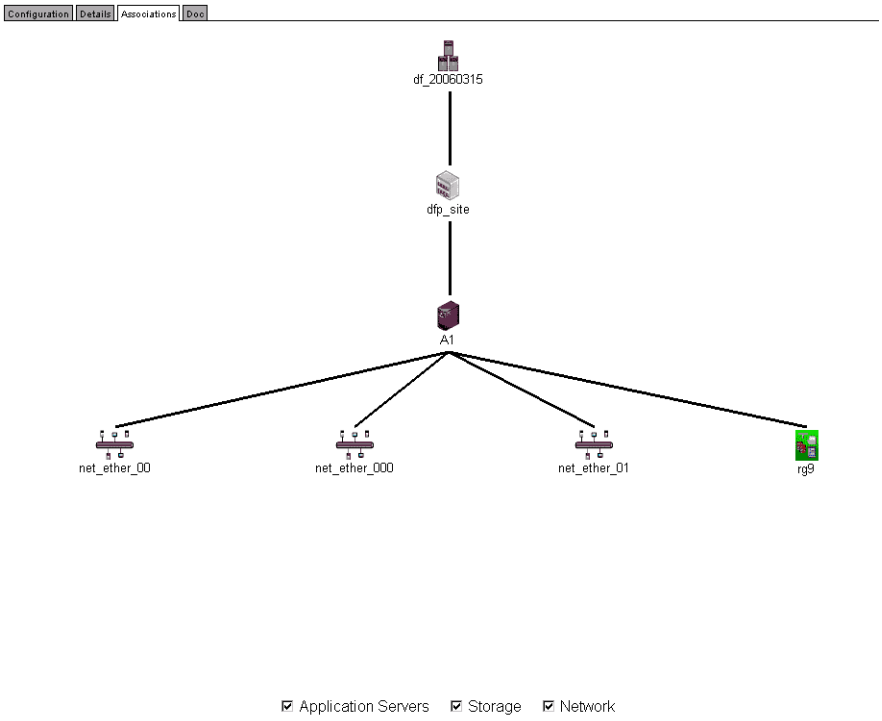
- The **N&N** view displays site- or node-centric associations related to the item selected in the navigation frame.
- The **RGs** view displays a resource-group-centric summary of where each resource group is online and any resource group dependencies.

When reviewing the **N&N** associations, if the graph becomes too complicated, you can remove the Application Servers, Storage, and the Networks from this view by de-selecting the check boxes at the bottom of the page. Similarly, when reviewing the **RGs** associations, if the graph becomes too complicated you can remove the Parent/Child, Online on Different Nodes, Online on Same Nodes, and Online on Same Site from the view by de-selecting the check boxes.

The following figures show examples of the associations displays:



WebSMIT: Resource Group Summary, Associations View



WebSMIT: Sites, Nodes and Networks Details, Associations View

Viewing Cluster Configuration Information in WebSMIT

The **Details** tab provides configuration information about the item selected in the Navigation frame. The following table shows the command used to retrieve the information, and the information displayed under the **Details** tab for each cluster component.

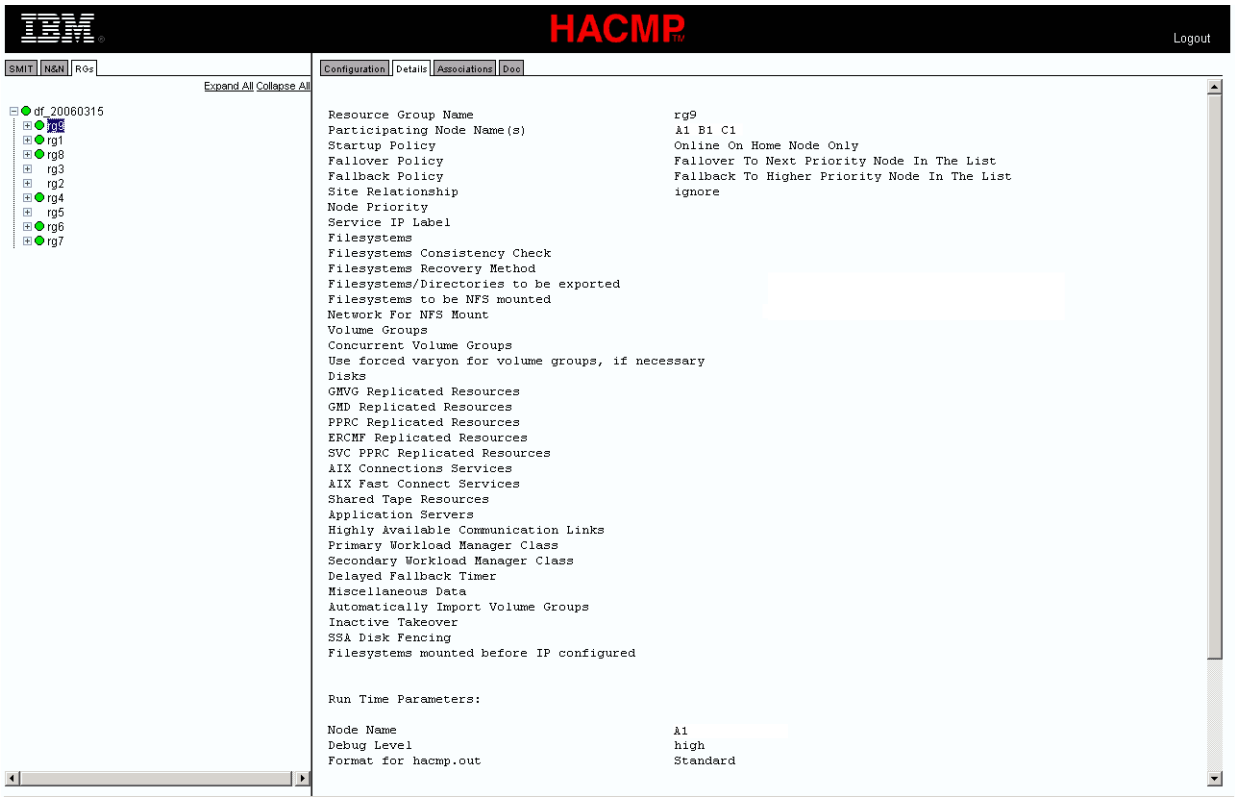
Component	Command	Details
Cluster	**cltopinfo**	Lists the cluster topology information.
Nodes	**cltopinfo**	Lists all node topology information.
Node	**cltopinfo -n**	Shows the configuration for the specified node and displays the status of the HACMP subsystems.
Network	**cltopinfo -w**	Shows all the networks configured in the cluster.
Resource Groups	**clshowres**	Shows the resources defined for all groups.
Resource Group	**clshowres -g**	Shows the resources defined to the selected group.

Volume Group	**lsvg <volume group name>**	Shows the status of the volume group.
Service IP	**cl_harvestIP_scripts -u <name>**	Lists the Service IP information.
Boot IP	**cltopinfo –i <name>**	Shows all interfaces configured in the cluster.
Application Server	**cllsserv -n <name>**	Lists application servers by name.

For more information on these commands, see the man page or the description in Appendix C: HACMP for AIX Commands.

The **Details** tab **c**ontains static information about the component selected in the left pane. This information is *not* automatically updated as changes are made to the cluster. To view updated information, click the item again.

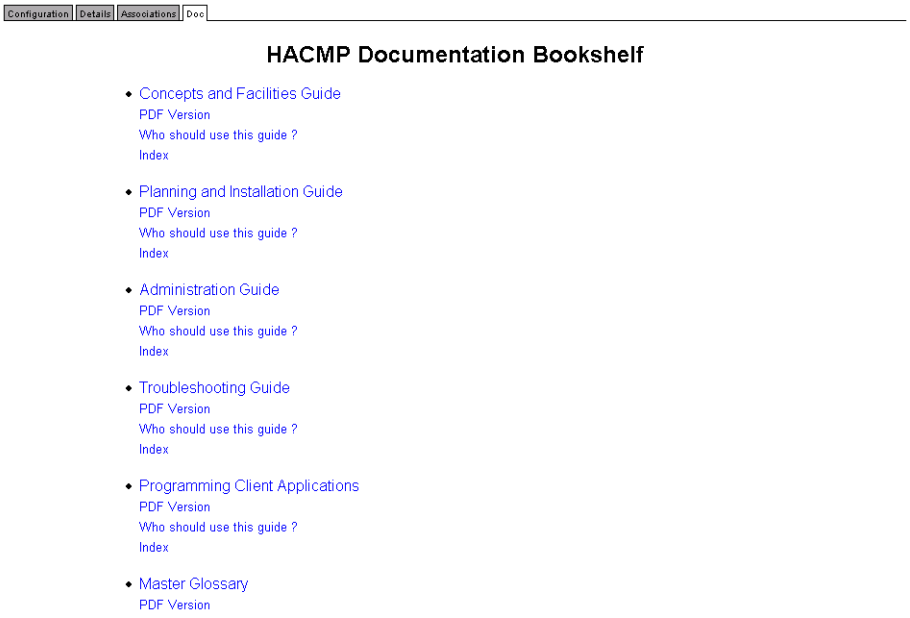
The following example shows the detail information shown for a specific resource group:



WebSMIT Resource Group Details Page

Viewing HACMP Documentation in WebSMIT

Through WebSMIT, you can access the HACMP documentation installed on your system. To view the WebSMIT bookshelf page, click the **Docs** tab in the **Activity** frame. The bookshelf page contains links to installed documentation only; WebSMIT notifies you of any documentation that is *not* installed.



WebSMIT Bookshelf Page

Chapter 3: Configuring an HACMP Cluster (Standard)

This chapter describes how to configure an HACMP cluster using the SMIT **Initialization and Standard Configuration** path.

Have your planning worksheets ready to help you through the configuration process. See the *Planning Guide* for details if you have not completed this step.

The main sections in this chapter include:

- Overview
- Configuring a Two-Node Cluster, or Using Smart Assists
- Defining HACMP Cluster Topology (Standard)
- Configuring HACMP Resources (Standard)
- Configuring HACMP Resource Groups (Standard)
- Configuring Resources in Resource Groups (Standard)
- Verifying and Synchronizing the Standard Configuration
- Viewing the HACMP Configuration
- Additional Configuration Tasks
- Testing Your Configuration.

Overview

Using the options under the SMIT **Initialization and Standard Configuration** menu, you can add the basic components of a cluster to the HACMP Configuration Database (ODM) in a few steps. This HACMP configuration path significantly automates the discovery and selection of configuration information and chooses default behaviors.

If you are setting up a basic two-node cluster, use the Two-Node Cluster Configuration Assistant to simplify the process for configuring a two-node cluster. For more information, see the section on Using the Two-Node Cluster Configuration Assistant in the chapter on Creating a Basic HACMP Cluster in the *Installation Guide.*

You can also use the General Configuration Smart Assist to quickly set up your application. You are not limited to a two-node cluster with this Assist.

You can use either ASCII SMIT or WebSMIT to configure the cluster. For more information on WebSMIT, see Chapter 2: Administering a Cluster Using WebSMIT.

Prerequisite Tasks for Using the Standard Path

Before using the Standard Configuration path, HACMP must be installed on all the nodes, and connectivity must exist between the node where you are performing the configuration and all other nodes to be included in the cluster. That is, network interfaces must be both physically and logically configured (to AIX 5L) so that you can successfully communicate from one node to each of the other nodes. The HACMP discovery process runs on *all* server nodes, not just the local node.

Once you have configured and powered on all disks, communication devices, serial networks and also configured communication paths to other nodes in AIX 5L, HACMP automatically collects information about the physical and logical configuration and displays it in corresponding SMIT picklists, to aid you in the HACMP configuration process.

With the connectivity path established, HACMP can discover cluster information and you are able to access all of the nodes to perform any necessary AIX 5L administrative tasks. That is, you do not need to open additional windows or physically move to other nodes' consoles, and manually log in to each node individually. To ease this process, SMIT fastpaths to the relevant HACMP and/or AIX 5L SMIT screens on the remote nodes are available within the HACMP SMIT screen paths.

HACMP uses all interfaces defined for the connectivity paths to populate the picklists. If you do not want a particular interface to be used in the HACMP cluster, use the HACMP Extended Configuration to delete it from the HACMP cluster (this way, it will not be shown up in picklists and will not be available for selection).

By default, cluster heartbeats are sent through all discovered networks. The network is kept highly available once it has an HACMP IP label assigned and you synchronize the configuration.

Assumptions and Defaults for the Standard Path

HACMP makes some assumptions regarding the environment, such as assuming all network interfaces on a physical network belong to the same HACMP network. Using these assumptions, HACMP supplies or automatically configures intelligent and default parameters to its configuration process in SMIT. This helps to minimize the number of steps it takes to configure the cluster.

HACMP makes the following basic assumptions:

- Hostnames are used as node names. HACMP automatically configures and monitors all network interfaces that can send a **ping** command to another network interface. The network interfaces that can send a **ping** command to each other without going through a router are placed on the same logical network. HACMP names each logical network.
- HACMP uses IP aliasing as the *default* mechanism for binding a service IP label/address to a network interface. For more information, see the chapter on planning the cluster networks in the *Planning Guide.*

Note: If you cannot use IP aliases because of hardware restrictions, such as the limited number of subnets that are allocated for cluster utilization, you will need to use IP replacement to bind your IP labels to network interfaces. For instance, ATM network does not support IP aliasing. IP replacement can be configured only under the SMIT **Extended Configuration** path (where you disable IP aliasing).

- IP Address Takeover via IP Aliases is configured for any logical network capable of taking over a service IP label as an alias. Note, in the Extended Configuration path, you can also configure IP Address Takeover that uses the IP replacement mechanism for binding IP labels/addresses with network interfaces.
- You can configure the resource groups with any of the policies for startup, fallover, and fallback (without specifying fallback timer policies).
- You can configure the application server start and stop scripts, but will need to use the Extended Configuration path to configure multiple monitors to track the health of each application server.

Also, since you can add, change, or remove serial (non-IP) networks and devices using the Extended Configuration path, you must manually define which pair of end-points exist in the point-to-point network, before adding, changing or removing serial networks and devices.

To manually configure any part of the cluster, or to add more details or customization to the cluster configuration, use the SMIT **HACMP Extended Configuration** path. See Chapter 4: Configuring HACMP Cluster Topology and Resources (Extended) for information on those options.

Note: If you are using the Standard Configuration path and information that is required for configuration resides on remote nodes, HACMP automatically discovers the necessary cluster information for you.

Steps for Configuring a Cluster Using the Initialization and Standard Configuration Path

Here are the steps to configure the typical cluster components:

What You Do	Description
Step 1: Configure a basic two-node cluster, or a cluster with WebSphere, DB2 or Oracle	*(Optional)* Use the **Configuration Assistants** panel to configure any of the following: • Basic two-node cluster • Cluster with the WebSphere application • Cluster with the DB2 UDB instances • Cluster with an Oracle application server and or Oracle database instances • Cluster with another application.

What You Do	Description
Step 2: Configure the cluster topology	Identify the cluster nodes and establish communication paths between them using the **Add Nodes to an HACMP Cluster** menu options. Here you name the cluster and select the nodes (listed in **/etc/hosts**) either by their names or their IP addresses. This gives HACMP the base knowledge it needs to communicate with the nodes that are participating in the cluster. Once each node is properly identified and HACMP obtains information about working communications paths, HACMP automatically runs a discovery operation that identifies the basic components within the cluster. HACMP uses the discovered hostnames as the node names and adds them to the HACMP Configuration Database (HACMPnode ODM). HACMP also automatically adds the networks and the associated interfaces that share physical connectivity with two or more nodes in the cluster to the HACMP Configuration Database (HACMPnetwork and HACMPadapter ODMs). Other resource information that HACMP discovers includes shared disk PVIDs and volume groups.
Step 3: Configure the cluster resources	Configure the resources to be made highly available. Use the **Configure Resources to Make Highly Available** menu to configure resources that are to be shared among the nodes in the cluster. You can configure these resources: • IP address/IP label • application server (a collection of start and stop scripts for the application that HACMP uses) • volume groups (shared and concurrent) • logical volumes • filesystems.
Step 4: Configure the resource groups	Use the **Configure HACMP Resource Groups** menu to create the resource groups you have planned for each set of related resources. You can configure startup, fallover and fallback policy for each resource group (without specifying fallback timer policies).
Step 5: Put the resources to be managed together into their respective resource groups	Use the **Configure HACMP Resource Groups > Change/Show Resources for a Resource Group (standard)** menu to assign resources to each resource group.

What You Do	Description
Step 6: Adjust log viewing and management	*(Optional)* Adjust log viewing and management (settings for the debug level and **hacmp.out** log file formatting options per node).
Step 7: Verify and synchronize the cluster configuration	Use the **Verify and Synchronize HACMP Configuration** menu to guarantee the desired configuration is feasible given the physical connections and devices, and ensure that all nodes in the cluster have the same view of the configuration.
Step 8: Display the cluster configuration	*(Optional)* Use the **Display HACMP Configuration** menu to view the cluster topology and resources configuration.
Step 9: Make further additions or adjustments to the cluster configuration	*(Optional)* You may want to use some options available on the **Extended Configuration** path. Such additions or adjustments include, for example: • Adding non-IP networks for heartbeating • Configuring and changing the distribution preference for service IP aliases • Adding other resources to the cluster, such as SNA communication interfaces and links or tape resources • Configuring resource group runtime policies, including Workload Manager • Adding resource group timers • Configuring dependencies between resource groups • Adding multiple application monitors for an application server • Configuring HACMP File Collections • Configuring cluster security • Customizing remote notifications (pager, SMS messages, and email) • Customizing cluster events • Configuring site policies.
Step 10: Test the cluster before it goes into the production environment	*(Recommended)* Use the **HACMP Cluster Test Tool** to test recovery procedures for the cluster.

Configuring a Two-Node Cluster, or Using Smart Assists

You can configure a basic two-node cluster with just a few configuration steps. For information, see the section on Using the Two-Node Cluster Configuration Assistant in the chapter on Creating a Basic HACMP Cluster in the *Installation Guide.*

If you are configuring a WebSphere, DB2 UDB or Oracle application, see the corresponding HACMP Smart Assist guide.

To configure other applications, you can use the General Configuration Smart Assist.

Limitations and Prerequisites

The initial requirements for using Smart Assists are:

- The application must be installed on all cluster nodes where you want to run it.
- The Smart Assist must be installed on all cluster nodes that run the application.

Configuring Applications with the General Configuration Smart Assist

To configure your installed application (other than DB2, WebSphere, or Oracle):

1. On a local node, enter `smitty hacmp`
2. Select **Initialization and Standard Configuration > Configuration Assistants > Make Applications Highly Available > Add an Application to the HACMP Configuration** and press Enter.

   If the cluster is not yet configured, you are directed to go to the **Configure HACMP Nodes and Cluster** SMIT panel. Here you need to list the communication paths to all nodes in the cluster. Then continue to the next step.
3. If the cluster is configured, SMIT displays a list of applications installed on this node. Select **Other Applications** and press Enter.
4. Select **General Application Smart Assist** and press Enter.
5. Enter values for the following fields on the **Add an Application to HACMP** panel:
   - **Application Server Name**
   - **Primary Node**
   - **Takeover Nodes**
   - **Application Server Start Script**
   - **Application Server Stop Script**
   - **Service IP Label**.
6. Press Enter after you have filled in the values. The configuration will be synchronized and verified automatically.
7. (Optional) Return to the panel **Make Applications Highly Available** to select **Test the HACMP Configuration** and press Enter.

   The Cluster Test Tool runs and displays results to the screen. If you get error messages, make the necessary corrections.

Defining HACMP Cluster Topology (Standard)

Complete the following procedures to define the cluster topology. You only need to perform these steps on one node. When you verify and synchronize the cluster topology, its definition is copied to the other nodes.

To configure the cluster topology:

1. Enter `smit hacmp`
2. In SMIT, select **Initialization and Standard Configuration > Configure an HACMP Cluster and Nodes** and press Enter.
3. Enter field values as follows:

Cluster Name	Enter an ASCII text string that identifies the cluster. The cluster name can include alphanumeric characters and underscores, but cannot have a leading numeric. Use no more than 32 characters. It can be different from the hostname. Do not use reserved names. For a list of reserved names see Chapter 7: Verifying and Synchronizing an HACMP Cluster.
New nodes (via selected communication paths	Enter (or add) one resolvable IP label (this may be the hostname), IP address, or Fully Qualified Domain Name for each new node in the cluster, separated by spaces. HACMP uses this path to initiate communication with the node. Example 1: *10.11.12.13 <space> NodeC.ibm.com*. Example 2: *NodeA<space>NodeB* (where these are hostnames.) The picklist displays the hostnames and/or addresses included in **/etc/hosts** that are not already HACMP-configured IP labels/addresses. You can add node names or IP addresses in any order.
Currently configured node(s)	If nodes are already configured, they are displayed here.

4. Press Enter. Once communication paths are established, HACMP runs the discovery operation and prints results to the SMIT panel.
5. Verify that the results are reasonable for your cluster.
6. Return to the top level HACMP SMIT panel to continue with the configuration.

Configuring HACMP Resources (Standard)

Using the Standard Configuration path, you can configure the types of resources that are most often included in HACMP clusters. You must first define resources that may be used by HACMP to the AIX 5L operating system on each node. Later you group together the associated resources in resource groups. You can add them all at once or add them separately, as you prefer.

This section shows how to configure the following components on all of the nodes defined to the cluster using a single network interface:

- Application servers (a collection of start and stop scripts that HACMP uses for the application).
- HACMP service IP labels/addresses. The service IP label/address is the IP label/address over which services are provided and which is kept highly available by HACMP.
- Shared volume groups, logical volumes, and filesystems.
- Concurrent volume groups, logical volumes, and filesystems.

Configuring Application Servers

An HACMP *application server* is a cluster resource used to control an application that must be kept highly available. It contains application start and stop scripts. Configuring an application server does the following:

- Associates a meaningful name with the server application. For example, you could give the application you are using with the HACMP software a name such as *apserv*. You then use this name to refer to the application server when you define it as a resource. When you set up the resource group that contains this resource, you define an application server as a resource.
- Points the cluster event scripts to the scripts that they call to start and stop the server application.
- Allows you to configure application monitoring for that application server. In HACMP 5.2 and up you can configure multiple application monitors for one application server. For more information see Steps for Configuring Multiple Application Monitors in Chapter 4: Configuring HACMP Cluster Topology and Resources (Extended).

Note that this section does not discuss how to write the start and stop scripts. See the vendor documentation for specific product information on starting and stopping a particular application.

Ensure that the server scripts exist on all nodes that participate as possible owners of the resource group where this application server resides.

To configure an application server on any cluster node:

1. Enter `smit hacmp`
2. In SMIT, select **Initialization and Standard Configuration > Configure Resources to Make Highly Available > Configure Application Servers > Add an Application Server** and press Enter.

3. SMIT displays the **Add an Application Server** panel. Enter field values as follows:

Server Name	Enter an ASCII text string that identifies the server. You will use this name to refer to the application server when you define resources during node configuration. The server name can include alphanumeric characters and underscores. Use a maximum of 64 characters.
Start Script	Enter the name of the script and its full pathname (followed by arguments) called by the cluster event scripts to start the application server. (Maximum 256 characters.) This script must be in the same location on each cluster node that might start the server. The contents of the script, however, may differ.
Stop Script	Enter the full pathname of the script called by the cluster event scripts to stop the server. (Maximum 256 characters.) This script must be in the same location on each cluster node that may start the server. The contents of the script, however, may differ.

4. Press Enter to add this information to the HACMP Configuration Database on the local node. Return to previous HACMP SMIT panels to perform other configuration tasks.

Configuring HACMP Service IP Labels/Addresses

A *service IP label/address* is used to establish communication between client nodes and the server node. Services, such as a database application, are provided using the connection made over the service IP label. This connection can be node-bound or taken over by multiple nodes. For the **Initialization and Standard Configuration** SMIT path, HACMP assumes that the connection will allow IP Address Takeover (IPAT) via IP Aliases (this is the default).

When you are using the standard configuration path, and add node names, IP labels/addresses, or hostnames to launch the initialization process, HACMP automatically discovers the networks for you.

The **/etc/hosts** file on all nodes must contain all IP labels and associated IP addresses that you want to discover.

Follow this procedure to define service IP labels for your cluster:

1. Enter `smit hacmp`
2. In SMIT, select **HACMP > Initialization and Standard Configuration** > **Configure Resources to Make Highly Available > Configure Service IP Labels/Addresses** and press Enter.

3. Fill in field values as follows:

IP Label/IP Address	Select from the picklist or enter the service IP label/address to be kept highly available. The name of the service IP label/address must be unique within the cluster and distinct from the volume group and resource group names; it should relate to the application it serves, as well as to any corresponding device, such as `websphere_service_address`.
Network Name	Enter the symbolic name of the HACMP network on which this Service IP label/address will be configured. If you leave the field blank, HACMP fills in this field automatically with the network type plus a number appended, starting with 1, for example, `netether1`.

4. Press Enter after filling in all required fields. HACMP now checks the validity of the IP interface configuration.
5. Repeat the previous steps until you have configured all service IP labels for each network, as needed.

Note: To control the placement of the service IP label aliases on the physical network interface cards on the cluster nodes, see the section in Chapter 4: Configuring HACMP Cluster Topology and Resources (Extended).

Configuring Volume Groups, Logical Volumes, and Filesystems as Cluster Shared Resources

You must define and properly configure volume groups, logical volumes and filesystems to AIX 5L, before using them as shared resources in an HACMP cluster. For information, see the relevant chapter in the *Installation Guide*, and Chapter 11: Managing Shared LVM Components in this Guide.

Configuring Concurrent Volume Groups, Logical Volumes, and Filesystems

These components must be defined to AIX 5L and properly configured, for use as shared resources. For information, see the relevant chapter in the *Installation Guide* and Chapter 12: Managing Shared LVM Components in a Concurrent Access Environment.

Configuring HACMP Resource Groups (Standard)

Refer to the *Concepts Guide* for an overview of resource groups you can configure in HACMP 5.4. Refer to the chapter on planning resource groups in the *Planning Guide* for further planning information. You should have your planning worksheets in hand.

Using the standard path, you can configure resource groups that use different startup, fallover, and fallback policies.

Once the resource groups are configured, if it seems necessary for handling certain applications, you can use the Extended Configuration path to change or refine the management policies of particular resource groups.

Configuring a resource group involves two phases:

1. Configuring the resource group name, startup, fallover and fallback policies, and the nodes that can own it (nodelist for a resource group)
2. Adding the resources and additional attributes to the resource group.

Refer to your planning worksheets as you name the groups and add the resources to each one.

Creating HACMP Resource Groups Using the Standard Path

To create a resource group:

1. Enter `smit hacmp`
2. In SMIT, select **Initialization and Standard Configuration** > **Configure HACMP Resource Groups > Add a Resource Group** and press Enter.
3. Enter information in the following fields:

Resource Group Name	Enter the name for this group. The name of the resource group must be unique within the cluster and distinct from the service IP label and volume group name. It is helpful to create the name related to the application it serves, as well as to any corresponding device, such as `websphere_service_address`. Use no more than 32 alphanumeric characters or underscores; do not use a leading numeric. Do not use reserved words. See List of Reserved Words. Duplicate entries are not allowed.
Participating Node Names	Enter the names of the nodes that can own or take over this resource group. Enter the node with the highest priority for ownership first, followed by the nodes with the lower priorities, in the desired order. Leave a space between node names, for example, `NodeA NodeB NodeX`.

Startup Policy

Select a value from the picklist that defines the startup policy of the resource group:

ONLINE ON HOME NODE ONLY. The resource group should be brought online *only* on its home (highest priority) node during the resource group startup. This requires the highest priority node to be available.

ONLINE ON FIRST AVAILABLE NODE. The resource group activates on the first node that becomes available.

ONLINE USING NODE DISTRIBUTION POLICY. If you select the node distribution policy, only one resource group is brought online on a node during startup.

Note: Rotating resource groups migrated from HACMP 5.1 now have this node-based distribution policy.

ONLINE ON ALL AVAILABLE NODES. The resource group is brought online on *all* nodes. This is equivalent to concurrent resource group behavior.

If you select this option for the resource group, ensure that resources in this group can be brought online on multiple nodes simultaneously.

Fallover policy

Select a value from the list that defines the fallover policy of the resource group:

FALLOVER TO NEXT PRIORITY NODE IN THE LIST. In the case of fallover, the resource group that is online on only one node at a time follows the default node priority order specified in the resource group's nodelist (it moves to the highest priority node currently available).

FALLOVER USING DYNAMIC NODE PRIORITY. If you select this option for the resource group (and Online on Home Node startup policy), you can choose one of the three predefined dynamic node priority policies. See Configuring Resource Group Runtime Policies in Chapter 5: Configuring HACMP Resource Groups (Extended).

BRING OFFLINE (ON ERROR NODE ONLY). Select this option to bring a resource group offline on a node during an error condition.

This option represents the behavior of a concurrent resource group and ensures that if a particular node fails, the resource group goes offline on that node only, but remains online on other nodes.

Selecting this option as the fallover preference when the startup preference is not Online On All Available Nodes may allow resources to become unavailable during error conditions. If you do so, HACMP issues an error.

Fallback policy	Select a value from the list that defines the fallback policy of the resource group: **NEVER FALLBACK**. A resource group does *not* fall back when a higher priority node joins the cluster. **FALLBACK TO HIGHER PRIORITY NODE IN THE LIST.** A resource group falls back when a higher priority node joins the cluster.

4. Press Enter.
5. Return to the **Add a Resource Group** panel to continue adding all the resource groups you have planned for the HACMP cluster.

Configuring Resources in Resource Groups (Standard)

After you have defined a resource group, assign resources to it. SMIT can list possible shared resources for the node if the node is powered on (helping you avoid configuration errors).

When you are adding or changing resources in a resource group, HACMP displays only valid choices for resources, based on the resource group management policies that you have selected.

Resource Group Configuration Considerations

Keep the following in mind as you prepare to define the resources in your resource group:

- You cannot configure a resource group until you have completed the information on the **Add a Resource Group** panel. If you need to do this, refer back to the instructions under Configuring HACMP Resource Groups (Standard).
- A resource group may include multiple service IP addresses. When a resource group configured with IPAT via IP Aliases is moved, all service labels in the resource group are moved as aliases to the available interfaces, according to the resource group management policies in HACMP.

  Also, you can specify the distribution preference for service IP labels. For more information, see Steps to Configure Distribution Preference for Service IP Label Aliases in the Chapter 4: Configuring HACMP Cluster Topology and Resources (Extended).

  For information on how HACMP handles the resource groups configured with IPAT via IP Aliases see Appendix B: Resource Group Behavior during Cluster Events.
- When you define a service IP label/address on a cluster node, the service label can be used in any non-concurrent resource group.
- IPAT function (both via IP Replacement and via IP Aliases) does not apply to concurrent resource groups (those with the startup policy Online on All Available Nodes).

Assigning Resources to Resource Groups (Standard)

To assign the resources for a resource group:

1. Enter `smit hacmp`

2. In SMIT, select **Initialization and Standard Configuration > Configure HACMP Resource Groups** > **Change/Show Resources for a Resource Group** and press Enter to display a list of defined resource groups.
3. Select the resource group you want to configure and press Enter. SMIT displays the panel that matches the type of resource group you selected, with the **Resource Group Name**, and **Participating Node Names (Default Node Priority)** fields filled in.

   Note: SMIT displays only valid choices for resources, depending on the type of resource group that you selected.

   If the participating nodes are powered on, you can press F4 to list the shared resources in the picklists. If a resource group/node relationship has not been defined, or if a node is not powered on, pressing F4 causes HACMP SMIT to display the appropriate warnings.
4. Enter the field values as follows:

Service IP Label/IP Addresses	This option appears only if you are adding resources to a non-concurrent resource group. List the service IP labels to be taken over when this resource group is taken over. See a picklist of valid IP labels. These include addresses that rotate or may be taken over.
Filesystems (empty is All for specified VGs)	This option appears only if you are adding resources to a non-concurrent resource group. If you leave the **Filesystems (empty is All for specified VGs)** field blank *and* specify the shared volume groups in the **Volume Groups** field below, all filesystems will be mounted in the volume group. If you leave the **Filesystems** field blank and do not specify the volume groups in the field below, no filesystems will be mounted. You may also select individual filesystems to include in the resource group. Press F4 to see a list of the filesystems. In this case only the specified filesystems will be mounted when the resource group is brought online.

Volume Groups

This option appears only if you are adding resources to a non-concurrent resource group.

Identify the shared volume groups that should be varied on when this resource group is acquired or taken over. Select the volume groups from the picklist or enter desired volume groups names in this field.

Press F4 for a list of all shared volume groups in the resource group *and* the volume groups that are currently available for import onto the resource group nodes.

Specify the shared volume groups in this field if you want to leave the field **Filesystems (empty is All for specified VGs)** blank *and* to mount all filesystems in the volume group. If you specify more than one volume group in this field, all filesystems in all specified volume groups are mounted. You cannot choose to mount all filesystems in one volume group and not to mount them in another.

For example, in a resource group with two volume groups, vg1 and vg2, if the **Filesystems (empty is All for specified VGs)** is left blank, all the filesystems in vg1 and vg2 are mounted when the resource group is brought up. However, if the **Filesystems (empty is All for specified VGs)** has only filesystems that are part of the vg1 volume group, none of the filesystems in vg2 are mounted, because they were not entered in the **Filesystems (empty is All for specified VGs)** field along with the filesystems from vg1.

If you have previously entered values in the **Filesystems** field, the appropriate volume groups are already known to HACMP.

Concurrent Volume Groups

This option appears only if you are adding resources to a non-concurrent resource group.

Identify the shared volume groups that can be accessed simultaneously by multiple nodes. Select the volume groups from the picklist, or enter desired volume groups names in this field.

If you previously requested that HACMP collect information about the appropriate volume groups, then the picklist displays a list of all existing concurrent capable volume groups that are currently available in the resource group, *and* concurrent capable volume groups available to be imported onto the nodes in the resource group.

Disk fencing is turned **on** by default.

Application Servers	Indicate the application servers to include in the resource group. The picklist displays a list of application servers.

Note: If you are configuring a resource group with the startup policy of Online on Home Node and the fallover policy Fallover Using Dynamic Node Priority, this SMIT panel displays the field where you can select which one of the three predefined dynamic node priority policies you want to use.

5. Press Enter to add the values to the HACMP Configuration Database.

Verifying and Synchronizing the Standard Configuration

After all resource groups have been configured, verify the cluster configuration on all nodes to ensure compatibility. If no errors are found, the configuration is then copied (synchronized) to each node of the cluster. If you synchronize from a node where Cluster Services are running, one or more resources may change state when the configuration changes take effect.

The Cluster Topology Summary

At the beginning of verification, before HACMP verifies the cluster topology, the Cluster Topology Summary is displayed listing any nodes, networks, network interfaces, and resource groups that are “unavailable” at the time that cluster verification is run. “Unavailable” refers to those that have failed and are considered offline by the Cluster Manager. These components are also listed in the **/var/hacmp/clverify/clverify.log** file.

The output from the verification is displayed in the SMIT Command Status window. If you receive error messages, make the necessary changes and run the verification procedure again.

The output may take one of the following forms:

- You may see warnings if the configuration has a limitation on its availability, for example, if only one interface per node per network is configured.
- Although no summary will be displayed to the user when no cluster topology components have failed, the **clverify.log** file displays the following:

  ```
  <DATE/TIME>  Verification detected that all cluster topology
  components are available.
  ```
- If cluster components are unavailable, the utility providing the list of failed components puts similar information in the log file.
- If the Cluster Manager is *not* running or is unavailable at the time when verification is run, only the **/var/hacmp/log/clutils.log**, file, *not* the user's display, is updated to include the following:

  ```
  Cluster Manager is unavailable on the local node. Failed components
  verification was not complete.
  ```
- If verification detects failed components, a wall message displays throughout all available cluster nodes, listing nodes, networks, network interfaces, and resource groups that are unavailable.

Procedure to Verify and Synchronize the HACMP Configuration

To verify and synchronize the cluster topology and resources configuration:

1. Enter `smit hacmp`
2. In SMIT, select **Initialization and Standard Configuration > Verify and Synchronize HACMP Configuration** and press Enter.
3. SMIT displays `Are you sure?` Press Enter again.

   SMIT runs the verification utility.

Viewing the HACMP Configuration

To display the HACMP cluster:

1. Enter `smit hacmp`
2. In SMIT, select **Initialization and Standard Configuration > Display HACMP Configuration** and press Enter.

   SMIT displays the current topology and resource information.

Additional Configuration Tasks

After you have finished the tasks on the **Initialization and Standard Configuration** menu and have synchronized the cluster configuration, consider further customizing the cluster. For example, you can:

- Define non-IP networks for heartbeats (highly recommended).
- Refine the distribution of service IP aliases placement on the nodes. For more information, see Steps to Configure Distribution Preference for Service IP Label Aliases in the Chapter 4: Configuring HACMP Cluster Topology and Resources (Extended).
- Configure dependencies between resource groups. Consider this step if you are planning to include multi-tiered applications in the cluster, where the startup of one application depends on the successful startup of another application.
- Refine resource groups behavior by specifying the delayed fallback timer, the settling time, and the node distribution policy.
- Configure a cluster snapshot so that it does not save log files.
- Configure multiple monitors for an application server, to monitor the health of your applications.
- Change runtime parameters and redirect log files for a node.
- Customize cluster events.
- Customize and configure different types of remote notification, such as pager, SMS messages, and email.
- Configure HACMP File Collections.
- Enable the cluster verification to run corrective actions.

See Chapter 4: Configuring HACMP Cluster Topology and Resources (Extended) to continue with these additional tasks.

Testing Your Configuration

After you configure a cluster, test it before using it in a production environment. For information about using the Cluster Test Tool to test your cluster, see Chapter 8: Testing an HACMP Cluster.

Chapter 4: Configuring HACMP Cluster Topology and Resources (Extended)

This chapter describes how to configure cluster topology and resources for an HACMP cluster using the SMIT **Extended Configuration** path.

The main sections in this chapter include:

- Understanding the Extended Configuration Options
- Discovering HACMP-Related Information
- Configuring Cluster Topology (Extended)
- Configuring HACMP Resources (Extended)
- Where You Go From Here.

Understanding the Extended Configuration Options

In order to configure the less common cluster elements, or if connectivity to each of the cluster nodes is unavailable, you can manually enter the information in a way similar to previous releases of the HACMP software.

When using the HACMP **Extended Configuration** SMIT paths, if any components are defined on remote nodes, you must manually initiate the discovery of cluster information. That is, discovery is optional (rather than automatic, as it is when using the **Initialization and Standard Configuration** SMIT path).

Using the options under the **Extended Configuration** menu, you can add the basic components of a cluster to the HACMP Configuration Database, as well as many additional types of resources. Use the **Extended Configuration** path to customize the cluster for all the components, policies, and options that are not included in the **Initialization and Standard Configuration** menus.

Note that certain options for configuring networks or IP labels are available only under the **Extended Configuration** path. In particular, make sure you use the **Extended Configuration** path if you plan to:

- Use IP Address Takeover via IP Replacement as a mechanism for binding IP labels/addresses to network interfaces. This option is available *only* under the **Extended Configuration** path in SMIT (configuring networks).
- Add or change an IP-based network.
- Add or change a non-IP-based network and devices.
- Configure persistent node IP labels, for cluster administrative purposes.
- Configure a distribution preference for service IP label aliases.

Note: You can use either ASCII SMIT or WebSMIT to configure a cluster. For more information on WebSMIT, see Chapter 2: Administering a Cluster Using WebSMIT.

Steps for Configuring an HACMP Cluster Using the Extended SMIT Menu

These are the basic steps to configure an HACMP cluster using the Extended SMIT menus.

What You Do	Description
Step 1: Run discovery	(*Optional)* Run discovery if you have already configured some or all of the cluster components. Running discovery retrieves current AIX 5L configuration information from all cluster nodes. This information is displayed in picklists to help you make accurate selections of existing components. HACMP informs you about which components have been discovered by the system. Predefined components (those that are supported but are not discovered) are also made available as selections. See Discovering HACMP-Related Information for details.

What You Do	Description
Step 2: Configure, change or customize the cluster topology	Under the **Extended Topology Configuration** menu, you can: • Identify the nodes and establish communication paths between them using the **Configure Nodes to an HACMP Cluster** menu options. Here you name the cluster and select the nodes (listed in **/etc/hosts**) either by their names or their IP addresses. This gives HACMP the information it needs to communicate with the nodes that are participating in the cluster. Once each of the nodes is properly identified and working communications paths exist, you can run discovery to identify the basic components within the cluster. The discovered hostnames are used as the node names and added to the HACMP Configuration Database (in particular, to HACMPnode ODM). The networks and the associated interfaces which share physical connectivity with two or more nodes in the cluster are automatically added to the HACMP Configuration Database (HACMPnetwork and HACMPadapter ODMs). Other discovered shared resource information includes PVIDs and volume groups. • (*Optional)* Configure, change, or show **sites**. • Configure, change, or show **predefined or discovered IP-based networks**, and **predefined or discovered serial devices**. • Configure, change, show and update with AIX 5L settings **HACMP Communication Interfaces and Devices**. You can configure either previously defined, or previously discovered communication interfaces and devices. • Configure, change, remove and show **Network Interface Modules (NIMs)**. • Configure, change, and show **Persistent Node IP Labels**.

What You Do	Description
Step 3: Configure or customize the Resources to be Made Highly Available	Use the **Extended Resource Configuration** menu to configure resources that are to be shared among the nodes in the cluster, such that if one component fails, another component will automatically take its place. You can configure the standard resources as well as several others: • IP labels • Application servers • Volume groups • Concurrent volume groups • Logical volumes • Filesystems • Application monitors • Tape resources • Communication adapters and links for the operating system • HACMP communication interfaces and links • Disk, volume group and filesystems methods for OEM disks, volumes and filesystems. In particular, you can configure Veritas volumes and filesystems to work in an HACMP cluster.
Step 4: Configure the resource groups	• Configure resource groups and define resource group policies and parameters • Configure resource group runtime policies. • Configure dependencies between resource groups (*optional*)
Step 5: Assign the resources that are to be managed together into resource groups	Place related resources into resource groups.

What You Do	Description
Step 6: Make any further additions or adjustments to the cluster configuration	(All *Optional*) • Configure cluster security. See Chapter 17: Managing Cluster Security. • Customize cluster events. See Chapter 6: Configuring Cluster Events. • Configure HACMP file collections. For more information, see Managing HACMP File Collections in Chapter 7: Verifying and Synchronizing an HACMP Cluster. • Configure performance tuning. See Chapter 1: Troubleshooting HACMP Clusters in the *Troubleshooting Guide*. • Configure a distribution preference for service IP label aliases. See Distribution Preference for Service IP Label Aliases: Overview in this chapter. • Customize remote notification. • Change attributes of nodes, communication interfaces and devices, networks, resources, or resource groups.
Step 7: Verify and synchronize the cluster configuration	Use the **Verify and Synchronize HACMP Configuration** menu to guarantee the desired configuration is feasible with the configured physical connections and devices, and ensure that all nodes in the cluster have the same view of the configuration.
Step 8: Display the cluster configuration	(*Optional*) View the cluster topology and resources configuration.
Step 9: Test cluster recovery procedures	*(Recommended)* Run automated or custom tests before putting the cluster in the production environment. Select **HACMP Cluster Test Tool** from the Extended Configuration SMIT path.

Discovering HACMP-Related Information

Running the cluster discovery process is optional in the **Extended Configuration** SMIT path.

After you have configured and powered on all disks, communication devices, point-to-point networks and configured communication paths to other nodes, HACMP can automatically collect this information and display it in corresponding SMIT picklists, to help you make accurate selections of existing components. HACMP informs you about which components have been discovered by the system. Pre-defined components (those that are supported but are not discovered) are also made available as selections.

Note: The discovery process runs on *all* nodes, not just on the local node.

To run the HACMP cluster discovery process, take the following steps:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Discover HACMP-related Information from Configured Nodes** and press Enter.
3. The software executes the discovery process.

Configuring Cluster Topology (Extended)

Complete the following procedures to define the cluster topology. You only need to perform these steps on one node. When you verify and synchronize the cluster topology, its definition is copied to the other node.

The Extended Topology Configuration panels include:

- Configuring an HACMP Cluster
- Resetting Cluster Tunables
- Configuring HACMP Nodes
- Defining HACMP Sites
- Configuring HACMP Networks and Heartbeat Paths
- Configuring Communication Interfaces/Devices to HACMP
- Configuring Heartbeating over Disk
- Configuring HACMP Persistent Node IP Labels/Addresses
- Configuring Node-Bound Service IP Labels
- Configuring HACMP Global Networks
- Configuring HACMP Network Modules
- Configuring Topology Services and Group Services Logs.

Configuring an HACMP Cluster

The only step necessary to configure the cluster is to assign the cluster name.

To assign a cluster name and configure a cluster:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Topology Configuration > Configure an HACMP Cluster** > **Add/Change/Show an HACMP Cluster** and press Enter.
3. Enter field values as follows:

Cluster Name	Enter an ASCII text string that identifies the cluster. The cluster name can include alphanumeric characters and underscores, but cannot have a leading numeric. Use no more than 32 characters. Do not use reserved names. For a list of reserved names see List of Reserved Words.

4. Press Enter.

5. Return to the **Extended Topology Configuration** SMIT panel.

Resetting Cluster Tunables

You can change the settings for a list of tunable values that were changed during the cluster maintenance and reset them to their default settings, or installation-time cluster settings.

Use this option to reset all the tunables (customizations) made to the cluster. Using this option returns all tunable values to their default values but does not change the cluster configuration. HACMP takes a snapshot file prior to resetting and informs you about the name and location of the snapshot file. You can choose to have HACMP synchronize the cluster when this operation is complete.

For instructions, see the Resetting HACMP Tunable Values section in Chapter 1: Troubleshooting HACMP Clusters in the *Troubleshooting Guide.* For a list of what tunable values will change, see the section on the List of Tunable Values provided in Chapter 1.

Configuring HACMP Nodes

To configure the cluster nodes:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Topology Configuration > Configure HACMP Nodes** > **Add a Node to the HACMP Cluster** and press Enter.
3. Enter field values as follows:

Node name	Enter a unique node name for the node. The name may be up to 32 characters in length. It is not required that the node name be the same as the host name of the node. You can enter one node name at a time, with up to 32 nodes in the cluster.
Communication Path to Node	Enter (or add) one resolvable IP Label (this may be the hostname), IP address, or Fully Qualified Domain Name for each new node in the cluster, separated by spaces. HACMP uses this path to initiate communication with the node. Example 1: *10.11.12.13 <space> NodeC.ibm.com*. Example 2: *NodeA<space>NodeB* (where these are hostnames.) Or, use the picklist to add the IP labels/addresses that are added to **/etc/hosts** but are not already configured in HACMP.

Once communication paths are established, HACMP adds a new node to the cluster.

Defining HACMP Sites

Site definitions are optional. They are supplied to provide easier integration with the HACMP/XD feature components: Metro Mirror (previously known as synchronous PPRC), GLVM, and HAGEO. Sites must also be defined if you want to use cross-site LVM mirroring.

If you define sites to be used in some other way, appropriate methods or customization must be provided to handle site operations. If sites are defined, site events run during **node_up** and **node_down** events. See Chapter 7: Planning Cluster Events, in the *Planning Guide* for more information.

For information and documentation for HACMP/XD, see About This Guide and the following URL:

http://www.rs6000.ibm.com/aix/library

For more information on cross-site LVM mirroring, see the *Planning Guide*.

If you are configuring sites, two sites must be configured and all nodes must belong to one of the two sites.

To add a site definition to an HACMP cluster:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Topology Configuration > Configure HACMP Sites** > **Add a Site Definition** and press Enter.
3. Enter the field values as follows. Also see the documentation for the product you are using that relies on the site definitions:

Site Name	Enter a name for this site using no more than 32 alphanumeric characters and underscores.
Site Nodes	Enter the names of the cluster nodes that belong to this site. Leave a space between names. A node can belong to only one site.
Dominance	Select **yes** or **no** to indicate whether the current site is primary or secondary. This field only applies to HACMP/XD for HAGEO configurations. It is ignored by other HACMP/XD software (Metro Mirror and GLVM).
Backup Communications	Select the type of backup communication for your cluster (you can also select **None**): • For HACMP/XD for HAGEO, select **DBFS** for telephone line, **SGN** for a Geo_Secondary network. • For HACMP/XD for GLVM, select **None**. • For HACMP/XD for Metro Mirror, select **None**.

4. Press Enter to add the site definition to the HACMP Configuration Database.
5. Repeat the steps to add the second site.

Configuring HACMP Networks and Heartbeat Paths

To avoid a single point of failure, the cluster should have more than one network. Often the cluster has both IP and non-IP based networks, which allows HACMP to use different heartbeat paths. Use the **Add a Network to the HACMP Cluster** SMIT panel to configure HACMP IP and point-to-point networks.

To speed up the configuration process, run discovery before configuring networks.

You can use any or all of these methods for heartbeat paths:

- Point-to-point networks
- IP-based networks, including heartbeating using IP aliases
- Heartbeating over disk. For information about configuring heartbeating over disk, see the section Configuring Heartbeating over Disk.

Note: When using an SP Switch network, configure an additional network for HACMP. If only one network is configured, HACMP issues errors during the cluster verification.

If you need to configure networks in HACMP for using them with in a cluster with sites that has HACMP/XD for GLVM installed, see the *HACMP/XD for GLVM Planning and Administration Guide* for the descriptions of XD_data, XD_ip and XD_rs232 networks.

Configuring IP-Based Networks

To configure IP-based networks:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Topology Configuration > Configure HACMP Networks > Add a Network to the HACMP Cluster** and press Enter.
3. Select the type of network to configure.
4. Enter the information as follows:

Network Name	If you do not enter a name, HACMP will give the network a default network name made up of the type of network with a number appended (for example, `ether1`). If you change the name for this network, use no more than 32 alphanumeric characters and underscores.
Network Type	This field is filled in depending on the type of network you selected.
Netmask	The netmask, for example, 255.255.255.0.

Enable IP Takeover via IP Aliases	The default is **True**. If the network does not support IP aliases, then IPAT via IP Replacement will be used. IP Replacement is the mechanism whereby one IP address is removed from an interface, and another IP address is added to that interface. If you want to use IP Replacement on a network that does support aliases, change the default to **False**.
IP Address Offset for Heartbeating over IP Aliases	Leave this field empty if you want HACMP to use the default heartbeating method. Enter a starting address for the heartbeat-over-alias network in this field. For example, `192.168.100.1`. You must include the starting address, not just the base network. Do not type in simply `192.168.100` or `192.168.100.0`, for example. The network you choose must be on a subnet that is not used by any other network in your physical network setup, and you must have enough available subnets above the one you type in for N networks, where N is the number of interfaces that each node has on the network. Using the example, here, you should have `192.168.100` and `192.168.101` subnets available for an HACMP network that has two interfaces on each node. HACMP uses this address to automatically generate IP addresses for heartbeating, for each boot interface in the configuration. This address range must be unique and must not conflict with any other subnets on the network. Refer to the section Heartbeating Over IP Aliases in Chapter 3: Planning Cluster Network Connectivity in the *Planning Guide* and your planning worksheet for more information on selecting a base address for use by Heartbeating over IP Aliases.

5. Press Enter to configure this network.
6. Repeat the operation to configure more networks.

Configuring IP Address Takeover via IP Replacement

If you do not have extra subnets to use in the HACMP cluster, you may need to configure IPAT via IP Replacement for the HACMP cluster network.

Note: IPAT via IP Aliases is the default method for binding an IP label to a network interface, and for ensuring the IP label recovery. IPAT via IP Aliases saves hardware, but requires multiple subnets. For general information about IPAT and the differences between the two IPAT methods, see the *Concepts and Facilities Guide*. For planning IPAT methods, see the *Planning Guide.*

To configure IPAT via IP Replacement:

1. In the **Add a Service IP Label/Address** SMIT panel, specify that the IP label that you add as a resource to a resource group is **Configurable on Multiple Nodes**.
2. In the same panel, configure hardware address takeover (HWAT) by specifying the **Alternate Hardware Address to Accompany IP Label/Address**. For instructions, see the section Configuring Service IP Labels/Addresses later in this chapter.

   Note: Do not use HWAT with gigabit Ethernet adapters (network interface cards) that support flow control. Network problems may occur after an HACMP fallover.

3. In the **Add a Network to the HACMP Cluster** panel, specify **False** in the **Enable IP Takeover via IP Aliases SMIT** field. For instructions, see the section Configuring Communication Interfaces/Devices to HACMP in this chapter.

Configuring Heartbeating over IP Aliases

You can configure heartbeating over IP Aliases to establish IP-based heartbeating rings over IP Aliases to run over your existing cluster networks. Heartbeating over IP Aliases supports either IPAT via IP Aliases or IPAT via IP Replacement. The type of IPAT configured determines how HACMP handles the service label:

IPAT via IP Aliases	The service label, as well as the heartbeat alias, is aliased onto the interface.
IPAT via IP Replacement	The service label is swapped with the interface IP address, not the heartbeating alias.

Note: HACMP removes the aliases from the interfaces at shutdown. It creates the aliases again when the network becomes operational. The **hacmp.out** file records these changes.

To configure heartbeating over IP Aliases, you specify an IP address offset when configuring an interface. Make sure that this address does not conflict with addresses configured on your network. For information about configuring an interface, see the section Configuring Communication Interfaces/Devices to HACMP.

Verifying Configuration for Heartbeating over IP Aliases

The HACMP cluster verification ensures that:

- The configuration is valid for the address range.
- All interfaces are the same type (for example, Ethernet) and have the same subnet mask.
- The offset address allows sufficient addresses and subnets on the network.

Configuring an Application Service Interface

If you already have an active application that is active and using a particular IP Address as a base address on network interface, you can configure this service IP label in HACMP without disrupting your application. The following steps guide you through configuring your application service IP label in HACMP in order not to disrupt your application:

1. Configure an HACMP cluster
2. Configure HACMP nodes
3. Configure HACMP networks
4. Run Discovery.
5. Configure HACMP communication interfaces/devices
6. Run verification and synchronization to propagate your configuration to all the nodes.
7. For each node that has an application using a particular IP Address:
   a. For each IP Address used as a base address on network interface on that node, decide on a Boot_IP_Address. For more information, see the *Planning Guide*.
   b. Run the sample utility **clchipdev** (described below):

   `/usr/es/sbin/cluster/samples/appsvclabel/clchipdev:`

The sample utility **clchipdev** helps configure an application service interface correctly in HACMP when you have an active application that is using a particular IP Address as a base address on network interface before starting HACMP.

```
clchdev -n NODE -w network_name -a 'App_IP_Address=Boot_IP_Address ...'
[-d alternate_HW_addr]
```

Where:

- NODE is the nodename.
- network_name is the name of the network that contains this service interface.
- App_IP_Address is the IP Address currently in use by the application (and currently configured in the CuAt as the base address for the given interface).
- Boot_IP_Address is the IP Address that is to be used as the new base (boot) address.
- Alternate_HW_address is an optional parameter for Hardware address (optionally used when configuring a service address in HA in IPAT networks).

For example, suppose NodeA has an IP Address 10.10.10.1 that is being used to make an application highly available. You would use the following steps:

1. Run the sample utility **clchipdev**.

   `clchdev -n NodeA -w net_ip -a '10.10.10.1=192.3.42.1'.`
   The sample utility performs the following:
   - Performs **rsh** to NodeA and determines the network interface for which 10.10.10.1 is currently configured as the base address.
   - Determines the network interface to be en0.
   - Determines the network type as defined in HACMPnetwork ODM, using the network name.

- Runs:

  ```
  chdev -l en0 -a netaddr=192.3.42.1 -P
  ```

  This changes the CuAt on that node to use the new Boot_IP_Address as the base address.
- Replaces 10.10.10.1 in HACMPadapter ODM with 192.3.42.1.
- Configures to HACMP the IP Address 10.10.10.1 as a service IP address.

2. Add this service IP label to a resource group.
3. Run verification and synchronization.

Configuring Point-to-Point Networks to HACMP

To configure a point-to-point network:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Topology Configuration > Configure HACMP Networks > Add a Network to the HACMP Cluster** and press Enter.

   SMIT displays a list of network types.
3. Select the type of network to configure.
4. Fill in the fields on the **Add a non IP-based Network** panel as follows:

Network Name	Name the network, using no more than 32 alphanumeric characters and underscores; do not begin the name with a numeric. Do not use reserved names. For a list of reserved names, see List of Reserved Words.
Network Type	Valid types are **RS232**, **tmssa**, **tmscsi**, **diskhb**.

Note: The volume groups associated with the disks used for disk heartbeating (disk heartbeating networks) do not have to be defined as resources within an HACMP resource group. In other words, an enhanced concurrent volume group associated with the disk that enables heartbeating does not have to belong to any resource group in HACMP.

5. Press Enter to configure this network.
6. Repeat the operation to configure more networks.

Configuring Communication Interfaces/Devices to HACMP

When you are configuring these HACMP components, you can have three different scenarios:

- Communication interfaces and devices are already configured to AIX 5L, and you have run the HACMP discovery process to add them to HACMP picklists to aid in the HACMP configuration process. See Configuring Discovered Communication Interfaces to HACMP and Configuring Discovered Communication Devices to HACMP.

- Communication interfaces and devices are already configured to AIX 5L, and need to be configured to HACMP (no discovery was run). See Configuring Predefined Communication Interfaces to HACMP and Configuring Predefined Communication Devices to HACMP.
- Network interfaces and serial devices need to be defined to AIX 5L before you can configure them in HACMP. In this case, HACMP SMIT provides you with links to the AIX 5L SMIT, where you can configure, change or delete communication interfaces/devices to the operating system, or update them with the AIX 5L settings without leaving the HACMP user interface.

  To configure network interfaces and serial devices to the AIX 5L operating system without leaving HACMP SMIT, use the **System Management (C-SPOC) > HACMP Communication Interface Management** SMIT path.

  For instructions, see the section Managing Communication Interfaces in HACMP in Chapter 13: Managing the Cluster Topology.

Configuring Discovered Communication Interfaces to HACMP

To add discovered communication interfaces to the HACMP cluster:

1. In SMIT, select **Extended Configuration > Extended Topology Configuration > Configure HACMP Communication Interfaces/Devices** > **Add Communication Interfaces/Devices** and press Enter. A panel appears that lets you add previously discovered, or previously defined network interfaces:

Add Discovered Communication Interfaces and Devices	Displays a list of interfaces and devices that HACMP has been able to determine as being already configured to AIX 5L on a node in the cluster.
Add Predefined Communication Interfaces and Devices	Displays a list of all communication interfaces and devices supported by HACMP.

2. Select the **Add Discovered Communication Interfaces and Devices** option. SMIT displays a list of interfaces and devices.
3. Select **Communication Interfaces** you want to configure. The panel **Select a Network Name** appears.
4. Select a network name.

   The panel **Select One or More Discovered Communication Interfaces to Add** appears. It displays a picklist that contains multiple communication interfaces, which when you select one or more, are added to the HACMP Configuration Database (ODM).

   HACMP either uses HACMP Configuration Database defaults, or automatically generates values, if you did not specifically define them earlier. For example, the **physical network name** is automatically generated by combining the string "**Net**" with the network type (for instance, **ether**) plus the next available integer, as in **NetEther3**.

Interfaces that are already added to the cluster are filtered from the picklist, as in the following example:

Node Name	Physical Network Name	Interface Name	IP Label	IP Address
All				
NodeA	NetEther3	en0	NodeA_en0	1.1.1.0
NodeA	NetEther2	en_1	NodeA_en1	1.1.1.1
NodeB	NetEther3	en0	NodeB_en0	1.1.2.0
NodeX	NetToken2	tr0	NodeX_tr0	1.1.1.3

5. Select **All**, one or more discovered communication interfaces to add. You have added either **All**, one or more of discovered communication interfaces to the operating system. (If you selected **All**, all discovered communication interfaces are added.)

Configuring Predefined Communication Interfaces to HACMP

The **Predefined Communication Interfaces** panel provides fields and picklists that enable you to choose configuration options quickly.

While choosing options, make sure that your choices do not conflict with the existing network topology. For example, if AIX 5L refers to a Token-Ring NIC (Network Interface Card), make sure that HACMP refers to the same type of network interface card (for example, not an Ethernet NIC).

To add predefined network interfaces to the HACMP cluster:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Topology Configuration > Configure HACMP Communication Interfaces/Devices** > **Add Communication Interfaces/Devices** and press Enter.

   A panel appears that lets you add previously discovered, or previously defined network interfaces:

Add Discovered Communication Interfaces and Devices	Displays a list of interfaces and devices, which HACMP has been able to determine as being already configured to the operating system on a node in the cluster.
Add Predefined Communication Interfaces and Devices	Displays a list of all communication interfaces and devices supported by HACMP.

3. Select the **Add Predefined Communication Interfaces and Devices** option. SMIT displays a list of communication interfaces and devices for the selected type.
4. Select **Communication Interfaces**. The **Select a Network Name** panel appears.
5. Select a network name. The **Add a Communication Interface** panel appears.

6. Fill in the fields as follows:

Node Name	The name of the node on which this network interface physically exists.
Network Name	A unique name for this logical network.
Network Interface	Enter the network interface associated with the communication interface (for example, `en0`).
IP Label/Address	The IP label/address associated with this communication interface that will be configured on the network interface when the node joins the cluster. The picklist filters out IP labels/addresses already configured to HACMP.
Network Type	The type of network media/protocol (e.g., `ethernet, token ring, fddi`, etc.) Select the type from the predefined list of network types.

7. Press Enter. You have added the communication interface(s) that were already predefined to AIX 5L to the HACMP cluster.

Configuring Discovered Communication Devices to HACMP

To configure discovered serial devices to the HACMP cluster:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Topology Configuration > Configure HACMP Communication Interfaces/Devices > Add Communication Interfaces/Devices** and press Enter. A panel appears that lets you add previously discovered, or previously defined network interfaces or devices:

Add Discovered Communication Interfaces and Devices	Displays a list of interfaces and devices that HACMP has been able to determine as being already configured to AIX 5L on a node in the cluster.
Add Predefined Communication Interfaces and Devices	Displays a list of all communication interfaces and devices supported by HACMP.

3. Select the **Add Discovered Communication Interfaces and Devices** option.
4. Select the **Communications Devices** type from the list.

   The panel **Select Point-to-Point Pair of Discovered Communication Devices to Add** appears. It displays a picklist that contains multiple communication devices, which when you select one or more, are added to the HACMP Configuration Database. Devices that are already added to the cluster are filtered from the picklist, as in the following example:

Node Name	Device	Device Path	PVID

```
NodeA       tty0        /dev/tty0
NodeA       tty1        /dev/tty1
NodeB       tty0        /dev/tty0
NodeB       tty1        /dev/tty1
NodeC       tty1        /devtty1
NodeC       tmssa0      /dev/tmssa0
NodeD       hdisk1      /dev/hdisk1   0002409f073
NodeE       hdisk1      /dev/hdisk1   0002409f073
```

5. Select *only* two devices in this panel. It is assumed that these devices are physically connected; you are responsible for making sure this is true.
6. Continue defining devices as needed.

Configuring Predefined Communication Devices to HACMP

To configure predefined communication devices for the cluster:

1. In SMIT, select **Extended Configuration > Extended Topology Configuration > Configure HACMP Communication Interfaces/Devices** > **Add Communication Interfaces/Devices** and press Enter. A panel appears that lets you add previously discovered, or previously defined network interfaces or devices:

Add Discovered Communication Interfaces and Devices	Displays a list of interfaces and devices that HACMP has been able to determine as being already configured to AIX 5L on a node in the cluster.
Add Predefined Communication Interfaces and Devices	Displays a list of all communication interfaces and devices supported by HACMP.

2. Select the **Add Predefined Communication Interfaces and Devices** option.
3. Select the **Communications Devices** type from the list.
   SMIT displays the **Add a Communications Device** panel.
4. Select the non IP-based network to which you want to add the devices.
5. Enter the field values as follows:

Node Name	The node name for the serial device.

Device Name	A device file name. RS232 serial devices must have the device file name **/dev/tty*n***. Target mode SCSI serial devices must have the device file name **/dev/tmscsi*n***. Target mode SSA devices must have the device file name **/dev/tmssa*n*.** For disk heartbeating, any disk device in an enhanced concurrent volume group is supported. It could be an **hdisk** or **vpath**, for example, **/dev/hdisk*n*** (*n* is the number assigned in each device file name).
Device Path	For an RS232, for example, **/dev/tty0**
Network Type	This field is automatically filled in (`RS232, tmssa, tmscsi, or diskhb`) when you enter the device name.
Network Name	This field is automatically filled in.

6. Press Enter after filling in all required fields. HACMP now checks the validity of the device configuration. You may receive warnings if a node cannot be reached.
7. Repeat until each node has all appropriate communication devices defined.

Configuring Heartbeating over Disk

You can configure disk heartbeating over any shared disk that is part of an enhanced concurrent mode volume group. RSCT passes topology messages between two nodes over the shared disk. Heartbeating networks contain:

- Two nodes
- An enhanced concurrent mode disk that participates in only one heartbeating network.

The following considerations apply:

- Use the **filemon** command to check the disk load. If the disk is heavily used, it may be necessary to change the tuning parameters for the network to allow more missed heartbeats. Disk heartbeating networks use the same tuning parameters as RS232 networks.
- For troubleshooting purposes, you or a third-party system administrator assisting you with cluster support may optionally reset the HACMP tunable values (such as aliases for heartbeating, or the tuning parameters for the network module) to their installation-time defaults. For more information, see the Resetting HACMP Tunable Values section in Chapter 1: Troubleshooting HACMP Clusters in the *Troubleshooting Guide.*

To configure a heartbeating over disk network:

1. Ensure that the disk is defined in an enhanced concurrent mode volume group.
2. Create a **diskhb** network.

   For information about configuring a **diskhb** network, see the section Configuring Point-to-Point Networks to HACMP.
3. Add disk-node pairs to the network.

   Ensure that each node is paired with the same disk, as identified by the device name, for example `hdisk1`, and PVID.

A device available as a physical volume to the Logical Volume Manager (LVM) is available for disk heartbeating, such as an `hdisk` or `vpath` disk. The **lspv** command displays a list of these devices.

For information about adding communication interfaces to a disk heartbeating network, see the sections Configuring Discovered Communication Interfaces to HACMP and Configuring Discovered Communication Devices to HACMP.

Verifying Disk Heartbeating Configuration

The HACMP cluster verification ensures the following:

- Interface and network names are valid and unique.
- Disk heartbeating devices use valid device names (**/dev/hdisk#**, **/dev/vpath#**).
- Disk heartbeating network devices are included in enhanced concurrent mode volume groups.
- Each heartbeating network has:
  - Two nodes with different names
  - Matching PVIDs for the node-hdisk pairs defined on each disk heartbeating network.

Using Disk Heartbeating Networks for Fast Failure Detection

HACMP 5.4 reduces the time it takes for a node failure to be realized throughout the cluster. When a node fails, HACMP uses disk heartbeating to place a departing message on the shared disk so neighboring nodes are aware of the node failure *within one heartbeat period* (hbrate). Topology Services then distributes the information about the node failure throughout the cluster nodes and a Topology Services daemon on each node sends a **node_down** event to any concerned client node.

You can turn on fast method of node failure detection when you change the NIM values for disk heartbeating networks. See Reducing the Node Failure Detection Rate: Enabling Fast Detection for Node Failures in Chapter 13: Managing the Cluster Topology.

Using Disk Heartbeating Networks for Detecting Failed Disk Enclosures

In addition to providing a non-IP network to help ensure high availability, disk heartbeating networks can be used to detect a failure of a disk enclosure. To use this function, configure a disk heartbeating network for at least one disk in each disk enclosure.

To let HACMP detect a failed disk enclosure:

1. Configure a disk heartbeating network for a disk in the specified enclosure.

   For information about configuring a disk heartbeating network, see the section Configuring Heartbeating over Disk.

2. Create a pre- or post-event, or a notification method, to determine the action to be taken in response to a failure of the disk heartbeating network. A failure of the disk enclosure is seen as a failure of the disk heartbeating network.

Configuring HACMP Persistent Node IP Labels/Addresses

A persistent node IP label is an IP alias that can be assigned to a network for a specified node. A persistent node IP label is a label that:

- Always stays on the same node (is node-bound)
- Co-exists with other IP labels present on an interface
- Does not require installing an additional physical interface on that node
- *Is not* part of any resource group.

Assigning a persistent node IP label for a network on a node allows you to have a node-bound address on a cluster network that you can use for administrative purposes to access a specific node in the cluster.

Prerequisites and Notes

If you are using persistent node IP Labels/Addresses, note the following issues:

- You can define only one persistent IP label on each node per cluster network.
- Persistent IP labels become available at a node's boot time.
- On a non-aliased network, a persistent label may be placed on the same subnet as the service labels, or it may be placed on an entirely different subnet. However, the persistent label must be placed on a different subnet than all non-service or non-boot IP labels (such as, "backup" IP labels) on the network.
- On an aliased network, a persistent label may be placed on the same subnet as the aliased service label, or it may be configured on an entirely different subnet. However, it must be placed on a different subnet than all boot IP labels on the network.
- Once a persistent IP label is configured for a network interface on a particular network on a particular node, it becomes available on that node on a boot interface at operating system boot time and remains configured on that network when HACMP is shut down on that node.
- You can remove a persistent IP label from the cluster configuration using the **Delete a Persistent Node IP Label/Address** SMIT panel. However, after the persistent IP label has been removed from the cluster configuration, it is not automatically deleted from the interface on which it was aliased. In order to completely remove the persistent IP label from the node, you should manually remove the alias with the **ifconfig delete** command or reboot the cluster node.
- Configure persistent node IP labels individually on each node. You cannot use the HACMP discovery process for this task.

To add persistent node IP labels:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Topology Configuration > Configure HACMP Persistent Node IP Labels/Addresses** > **Add a Persistent Node IP Label** and press Enter.
3. Enter the field values as follows:

Node Name	The name of the node on which the IP label/address will be bound.
Network Name	The name of the network on which the IP label/address will be bound.

Node IP Label/Address The IP label/address to keep bound to the specified node.

4. Press Enter.

To change or show persistent node IP labels, use the **Change/Show a Persistent Node IP label** SMIT menu. To delete them, use the **Delete a Persistent Node IP label** menu.

Configuring Node-Bound Service IP Labels

A *node-bound service IP label* is a specific type of service IP label configured on a non-aliased network.

In the case of node-bound service IP labels, IP aliases are not used, and therefore, to have a node-bound service IP label, a network should be first configured to use IPAT via IP Replacement.

These IP labels do not "float" with a resource group, but they are kept highly available on the node to which they are assigned. Node-bound service IP labels could be useful for the following purposes:

- A concurrent resource group could have node-bound service IP addresses configured for it. This way, a "load balancer" type of application (or another application, in general) configured in front of the HACMP cluster could forward requests to each instance of the concurrent resource group via specifically assigned node-bound service IP addresses on each of the nodes in the resource group nodelist.
- A node-bound service IP label could also be used for administrative purposes. Some of its functions can, of course, be achieved by using other capabilities in HACMP (such as using persistent service IP labels).
- Node-bound service IP labels are required for configuring clusters in HACMP/XD for HAGEO. For more information, see the HACMP/XD for HAGEO documentation.

To configure a node-bound service IP label:

1. Add an IP-based network to the HACMP cluster, use the procedure in the section Configuring IP-Based Networks.
2. Disable the option for IP Address Takeover via IP Aliases for this network. Use the procedure in the section Configuring IP Address Takeover via IP Replacement.
3. Add a communication interface to the network. Use the procedures in the section Configuring Discovered Communication Interfaces to HACMP, or in the section Configuring Discovered Communication Devices to HACMP.
4. Add a service IP label/address that is bound to a single node. Use the procedure in the section Configuring Service IP Labels/Addresses and select an option **Bound to a Single Node**.

Configuring HACMP Global Networks

This section describes global networks and steps to configure them.

HACMP Global Networks: Overview

In order to reduce the possibility of a partitioned HACMP cluster, you should configure multiple heartbeat paths between cluster nodes. Even if an IP network is not used for IP address takeover (IPAT), it should still be defined to HACMP for use by heartbeat. This reduces the chance that the loss of any single network will result in a partitioned cluster (and subsequent shutdown of one side of the partition).

When defining IP networks for heartbeating only (for example, the Administrative Ethernet on an SP), it is possible to combine multiple individual networks into a larger *global* network. This allows heartbeating across interfaces on multiple subnets.

Interfaces on each subnet are added to a single HACMP network (either manually or using the HACMP discovery process), and then the individual networks are added to the global network definition. HACMP treats all interfaces on the combined global network as a single network and ignores the subnet boundaries of the individual networks.

Networks combined in a global network therefore cannot be used for IP address recovery (that is, the IP labels from such networks should not be included in resource groups).

You define global networks by assigning a character string name to each HACMP network that you want to include as a member of the global network. All members of a global network must be of the same type (all Ethernet, for example).

Steps for Configuring HACMP Global Networks

To configure global networks, complete the following steps:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Topology Configuration > Configure HACMP Global Networks** and press Enter. SMIT displays a picklist of defined HACMP networks.
3. Select one of these networks. SMIT displays the **Change/Show a Global Network** panel. The name of the network you selected is entered as the local network name.
4. Enter the name of the global network (character string).
5. Repeat these steps to define all the HACMP networks to be included in each global network you want to define.

Configuring HACMP Network Modules

Using HACMP SMIT, you can add, change, remove or list existing network interface modules (NIMs).

For information on listing existing NIMs, or adding and removing NIMs, see Changing the Configuration of a Network Module in Chapter 13: Managing the Cluster Topology.

To add a network notifies module:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Topology Configuration> Configure HACMP Network Modules** > **Add a Network Module** and press Enter.

3. Enter field values as follows:

Network Module Name	Name of network type, for example, ether.
Description	For example, Ethernet Protocol
Address Type	Select an option: **Device** or **Address**. The **Address** option specifies that the network interface associated with this network module uses an IP-type address. The **Device** option specifies that the network interface associated with this network module uses a device file.
Path	Specifies the path to the network executable file.
Parameters	Specifies the parameters passed to the NIM executable. For the RS232 NIM, this field specifies the baud rate. Allowable values are 38400 (the default), 19200 and 9600.
Grace Period	The current setting is the default for the network module selected. This is the time period during which, after a network failure was detected, further network failures of the same type would be ignored. This is 60 seconds for all networks except ATM and Token Ring, which are 90 seconds.
Entry Type	This field specifies the type of the network interface - either a network interface card (for a NIM specific to a network interface card), or a network interface type (for a NIM to use with a specific type of network device).
Next Generic Type	This field specifies the next type of NIM to try to use if a more suitable NIM cannot be found.
Next Generic Name	This field specifies the next generic NIM to try to use if a more suitable NIM cannot be found.
Supports Source Routing	Set this field to **true** if this network supports IP loose source routing.
Failure Cycle	The current setting is the default for the network module selected. (Default for Ethernet is 10). This is the number of successive heartbeats that can be missed before the interface is considered to have failed. You can enter a number from 1 to 75.
Interval between Heartbeats (seconds)	The current setting is the default for the network module selected and is a heartbeat rate. This parameter tunes the interval (in seconds) between heartbeats for the selected network module. You can enter a number from less than 1 to 5.

After the command completes, a panel appears that shows the current settings for the specified network module.

Configuring Topology Services and Group Services Logs

You can change the settings for the length of the Topology and Group services logs. However, the default settings are highly recommended. The SMIT panel contains entries for heartbeat settings, but these are not adjustable.

Note: You can change the HACMP network module settings. See the section Configuring HACMP Network Modules.

To configure Topology Services and Group Services logs:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Topology Configuration > Configure Topology Services and Group Services > Change/Show Topology and Group Services Configuration** and press Enter.

   SMIT displays the **Change/Show Topology and Group Services Configuration** panel.
3. Enter field values as follows:

   Topology Services log length (lines) The default is 5000. This is usually sufficient.

   Group Services log length (lines) The default is 5000. This is usually sufficient.
4. Press Enter if you make any changes to field values, and then return to the menu.

Showing HACMP Topology

To view the current HACMP topology configuration, select the **Show HACMP Topology** option from the **Extended Configuration** > **Extended Topology Configuration** menu.

Configuring HACMP Resources (Extended)

Once you have configured the cluster topology, continue setting up your cluster by configuring the resources that will be placed in the resource groups. You may have already used the HACMP Standard path menus to configure some resources and groups. Use the Extended menus to add the resources not available on the standard path, to make changes, or to add more extensive customization.

Using the **Extended Resources Configuration** path you can configure the following types of resources:

- Application server
- Service IP label
- Shared volume group
- Concurrent volume group
- Filesystem
- Application monitor(s)

- CS/AIX
- Fast Connect
- Tape drive.

Configuring Service IP Labels as HACMP Resources

You should record the service IP label configuration on the planning worksheets. See Chapter 3: Planning Cluster Network Connectivity in the *Planning Guide* for information on service IP labels that are included as resources in the resource groups.

For the initial configuration, follow the procedures described in this section.

Discovering IP Network Information (Optional)

When you are using the extended configuration path, you can choose to run the HACMP cluster information discovery process. If you choose to run discovery, all communication paths must be configured first. Then HACMP will discover nodes, networks, and communication interfaces and devices for you and show them in the SMIT picklists. If you choose not to run discovery, HACMP will only include in the picklist network information that is predefined in AIX 5L.

To run cluster discovery:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Discover HACMP-Related Information from Configured Nodes** and press Enter.

   HACMP retrieves current AIX 5L configuration information from all cluster nodes. This information is displayed in picklists to help you make accurate selections of existing components. HACMP informs you about which components have been discovered by the system. Predefined components (those that are supported but are not discovered) are also made available as selections in picklists.
3. Return to the **Extended Configuration** menu.

Configuring Service IP Labels/Addresses

To add service IP labels/addresses as resources to the resource group in your cluster:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration** > **Extended Resource Configuration > HACMP Extended Resources Configuration** > **Configure HACMP Service IP Labels/Addresses > Add a Service IP Label/Address** and press Enter. SMIT displays the following panel.

3. Select the type of service IP label you are configuring:

Configurable on multiple nodes	By default, a service IP label/address may be acquired by multiple nodes, although it will be configured on only one node at any given time. (This address is allowed to participate in IP Address Takeover, either via IP Aliases or via IP Replacement.) Choosing this option will maintain the Service IP label/address even when the node which currently owns the Service IP label/address fails.
Bound to a single node	Select this option to maintain this service IP label/address on the same node at all times. The service IP label/address will be kept highly available as long as this node has an active network interface available on this node on the associated network. However, if the node itself fails, or the node has no available network interfaces, the Service IP label/address will become unavailable. Note, if this option is selected, and IPAT via IP Aliases is disabled for the network of which this interface is a part, the service IP label/address *must* also be configured to the AIX 5L operating system (in the ODM) of the node on which it is to be maintained.

4. Fill in field values as follows:

IP Label/Address	Enter, or select from the picklist the IP label/address to be kept highly available.
Network Name	Enter the symbolic name of the HACMP network on which this Service IP label/address will be configured.
Alternate Hardware Address to Accompany IP Label/Address	*If you chose a node-bound IP label, this field will not display*. Enter a hardware address for the service IP label. The hardware address must be unique within the physical network. Enter a value in this field *only* if you are currently defining a service IP label *and* you want to use hardware address swapping. This facility is supported for Ethernet, Token-Ring, FDDI and ATM. It does not work with the SP Switch. *Do not enter* a value in this field if you are configuring the network for IPAT via IP Aliases. Note that the hardware address is 12 digits for Ethernet, Token-Ring and FDDI; and 14 digits for ATM.

Node	*If you chose an IP label configurable on multiple nodes, this field will not display*. The name of the node where this IP label/address should be bound. **Note**: If this option is selected, and IPAT via IP Aliases is disabled for the network where this interface resides, the service label/address *must* be configured to AIX 5L (in the CuAt ODM) of the node where it is to be maintained.
Associated Site	*If you chose a node-bound IP label, this field will not display*. **Ignore** is the default. The service IP label can be activated on any node regardless of site. Select a site name to associate with the service IP label. This service IP label can only be activated on nodes belonging to the associated site. *Site-specific service IP labels are only activated when their resource group is online primary on the specified site.*

5. Press Enter after filling in all required fields. HACMP now checks the validity of the IP label/address configuration.
6. Repeat the previous steps until you have configured all service IP labels/addresses for each network, as needed.

Distribution Preference for Service IP Label Aliases: Overview

You can configure a distribution preference for the service IP labels that are placed under HACMP control. HACMP lets you specify the distribution preference for the service IP label aliases. These are the service IP labels that are part of HACMP resource groups and that belong to IPAT via IP Aliasing networks.

A *distribution preference for service IP label aliases* is a network-wide attribute used to control the placement of the service IP label aliases on the physical network interface cards on the nodes in the cluster. Configuring a distribution preference for service IP label aliases does the following:

- Lets you customize the load balancing for service IP labels in the cluster.
- Enables HACMP to redistribute the alias service IP labels according to the preference you specify.
- Allows you to configure the type of distribution preference suitable for the VPN firewall external connectivity requirements. For more information, see the section *Planning for the VPN Firewall Network Configurations in HACMP* in the *Planning Guide.*
- The distribution preference is exercised as long as there are acceptable network interfaces available. HACMP always keeps service IP labels active, even if the preference cannot be satisfied.

Rules for the Distribution Preference for Service IP Label Aliases

The following rules apply to the distribution preference:

- You can specify the distribution preference for service IP labels that belong to IPAT via IP Aliases networks.
- If you do not specify any preference, HACMP by default distributes all service IP label aliases across all available boot interfaces on a network using the IPAT via IP Aliasing function. For more information on how the default method for service IP label distribution works, see Appendix B: Resource Group Behavior during Cluster Events.
- If there are insufficient network interface cards available to satisfy the preference that you have specified, HACMP allocates service IP label aliases to an active network interface card that may be hosting other IP labels.
- You can change the IP labels distribution preference dynamically: The new selection becomes active during subsequent cluster events. (HACMP does not require the currently active service IP labels to conform to the newly changed preference.)
- If you did not configure persistent labels, HACMP lets you select the Collocation with Persistent and Anti-Collocation with Persistent distribution preferences, but it issues a warning and uses the regular collocation or anti-collocation preferences by default.
- When a service IP label fails and another one is available on the same node, HACMP recovers the service IP label aliases by moving them to another NIC on the same node. During this event, the distribution preference that you specified remains in effect.
- You can view the distribution preference per network using the **cltopinfo** or the **cllsnw** commands.

Configuring Distribution Preference for Service IP Label Aliases

To specify a distribution preference for service IP label aliases:

1. *(Optional)*. Configure a persistent IP label for each cluster node on the specific network. For instructions, see the section Configuring HACMP Persistent Node IP Labels/Addresses.
2. Configure the service IP labels for the network.
3. Select the type of the distribution preference for the network. For a list of available distribution preferences, see the section below.

Types of Distribution for Service IP Label Aliases

You can specify in SMIT the following distribution preferences for the placement of service IP label aliases:

Type of distribution preference	Description
Anti-collocation	This is the default. HACMP distributes all service IP label aliases across all boot IP labels using a “least loaded” selection process.
Collocation	HACMP allocates all service IP label aliases on the same network interface card (NIC).

Type of distribution preference	Description
Anti-collocation with persistent	HACMP distributes all service IP label aliases across all active physical interfaces that are *not* hosting the persistent IP label. HACMP will place the service IP label alias on the interface that is hosting the persistent label only if no other network interface is available. If you did not configure persistent IP labels, HACMP lets you select the Anti-Collocation with Persistent distribution preference, but it issues a warning and uses the regular anti-collocation preference by default.
Collocation with persistent	All service IP label aliases are allocated on the same NIC that is hosting the persistent IP label. This option may be useful in VPN firewall configurations where only one interface is granted external connectivity and all IP labels (persistent and service) must be allocated on the same interface card. If you did not configure persistent IP labels, HACMP lets you select the Collocation with Persistent distribution preference, but it issues a warning and uses the regular collocation preference by default.

Steps to Configure Distribution Preference for Service IP Label Aliases

To configure a distribution preference for service IP label aliases on any cluster node:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Resource Configuration > HACMP Extended Resources Configuration > Configure Resource Distribution Preferences > Configure Service IP labels/addresses Distribution Preferences** and press Enter.

   HACMP displays a list of networks that use IPAT via IP Aliasing.
3. Select a network for which you want to specify the distribution preference.
4. SMIT displays the **Configure Resource Distribution Preferences** screen. Enter field values as follows:

Network Name	The field is filled in with the network for which you want to specify or change the distribution preference for service IP label aliases.

Distribution Preference From the picklist, select the distribution preference as follows:

- **Anti-collocation**. This is the default. HACMP distributes all service IP label aliases across all boot IP labels using a least loaded selection process.
- **Collocation**. HACMP allocates all service IP label aliases on the same network interface card (NIC).
- **Anti-collocation with persistent**. HACMP distributes all service IP label aliases across all active physical interfaces that are *not* hosting the persistent IP label. **Note**: HACMP allocates the service IP label alias on the interface that is hosting the persistent label *only* if no other interface is available.
- **Collocation with persistent**. All service IP label aliases are allocated on the same NIC that is hosting the persistent IP label. This option may be useful in firewall configurations where only one interface is granted external connectivity and all IP labels (persistent and service) must be allocated on the same interface card.

Note: If you did not configure persistent IP labels, HACMP lets you select the **Collocation with Persistent** and **Anti-Collocation with Persistent** distribution preferences but issues a warning and uses the regular collocation or anti-collocation preferences by default.

5. Press Enter to add this information to the HACMP Configuration Database on the local node. Return to previous HACMP SMIT screens to perform other configuration tasks.
6. To synchronize the cluster definition, go to the **Initialization and Standard Configuration** or **Extended Configuration** menu and select **Verification and Synchronization**. If the Cluster Manager is running on the local node, synchronizing the cluster resources triggers a dynamic reconfiguration event.

   See Synchronizing Cluster Resources in Chapter 14: Managing the Cluster Resources for more information.

Configuring HACMP Application Servers

An *application server* is a cluster component that is included in the resource group as a cluster resource, and that is used to control an application that must be kept highly available. An application server consists of application start and stop scripts. Configuring an application server does the following:

- Associates a meaningful name with the server application. For example, you could give the tax software a name such as *taxes*. You then use this name to refer to the application server when you define it as a resource. When you set up the resource group, you add an application server as a resource.
- Points the cluster event scripts to the scripts that they call to start and stop the server application.

- Allows you to then configure application monitoring for that application server.

Note that this section does not discuss how to write the start and stop scripts. See the vendor documentation for specific product information on starting and stopping a particular application.

Defining an HACMP Application Server

To configure an application server on any cluster node:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Resource Configuration > HACMP Extended Resources Configuration > Configure HACMP Applications > Configure HACMP Application Servers > Add an Application Server** and press Enter.

   SMIT displays the **Add an Application Server** panel.
3. Enter field values as follows:

Server Name	Enter an ASCII text string that identifies the server. You will use this name to refer to the application server when you define resources during node configuration. The server name can include alphabetic and numeric characters and underscores. Use no more than 64 characters.
Start Script	Enter the pathname of the script (followed by arguments) called by the cluster event scripts to start the application server. (Maximum 256 characters.) This script must be in the same location on each cluster node that might start the server. The contents of the script, however, may differ.
Stop Script	Enter the pathname of the script called by the cluster event scripts to stop the server. (Maximum 256 characters.) This script must be in the same location on each cluster node that may start the server. The contents of the script, however, may differ.

4. Press Enter to add this information to the HACMP Configuration Database on the local node. Return to previous HACMP SMIT panels to perform other configuration tasks.

Configuring Volume Groups, Logical Volumes, and Filesystems as Resources

You define volume groups, logical volumes, and filesystems in AIX 5L and then configure them as resources for HACMP. Plan and note them on the worksheets before configuring to HACMP. For more information, see Chapter 5 in the *Installation Guide* and Chapter 11: Managing Shared LVM Components in this guide.

Configuring Concurrent Volume Groups, Logical Volumes, and Filesystems as Resources

Concurrent volume groups, logical volumes, and filesystems must be defined in AIX 5L and then configured as resources for HACMP. They should be planned and noted on the worksheets before configuring to HACMP. See Chapter 5: Planning Shared LVM Components in the *Planning Guide* and Chapter 12: Managing Shared LVM Components in a Concurrent Access Environment for information in this guide.

Configuring Multiple Application Monitors

HACMP can monitor specified applications using application monitors. These application monitors can:

- Check if an application is running before HACMP starts it.
- Watch for the successful startup of the application.
- Check that the application runs successfully after the stabilization interval has passed.
- Monitor both the startup and the long-running process.
- Automatically take action to restart applications upon detecting process termination or other application failures.

In HACMP 5.2 and up, you can configure multiple application monitors and associate them with one or more application servers.

By supporting multiple monitors per application, HACMP can support more complex configurations. For example, you can configure one monitor for each instance of an Oracle parallel server in use. Or, you can configure a custom monitor to check the health of the database along with a process termination monitor to instantly detect termination of the database process.

Note: If a monitored application is under control of the system resource controller, ensure that `action:multi` are **-O** and **-Q**. The **-O** specifies that the subsystem is not restarted if it stops abnormally. The **-Q** specifies that multiple instances of the subsystem are not allowed to run at the same time. These values can be checked using the following command:

```
lssrc -Ss Subsystem | cut -d : -f 10,11
```

If the values are not **-O** and **-Q**, change them using the **chssys** command.

Process and Custom Monitoring

You can select either of two application monitoring methods:

- *Process application monitoring* detects the termination of one or more processes of an application, using RSCT Resource Monitoring and Control.
- *Custom application monitoring* checks the health of an application with a custom monitor method at user-specified polling intervals.

Process monitoring is easier to set up, as it uses the built-in monitoring capability provided by RSCT and requires no custom scripts. However, process monitoring may not be an appropriate option for all applications. Custom monitoring can monitor more subtle aspects of an application's performance and is more customizable, but it takes more planning, as you must create the custom scripts.

Fallover and Notify Actions

In both process and custom monitoring methods, when the monitor detects a problem, HACMP attempts to restart the application on the current node and continues the attempts until the specified retry count has been exhausted.

When an application cannot be restarted within the retry count, HACMP takes one of two actions, which you specified when configuring the application monitor:

- Choosing **fallover** causes the resource group containing the application to fall over to the node with the next highest priority according to the nodelist. (See Note on the Fallover Option and Resource Group Availability for more information.)
- Choosing **notify** causes HACMP to generate a **server_down** event, which informs the cluster of the failure.

Monitor Modes

When you configure process monitor(s) and custom monitor(s) for the application server, you can also specify the mode in which the application monitor is used:

- *Startup Monitoring Mode*. In this mode, the monitor checks the application server's successful startup within the specified stabilization interval and exits after the stabilization period expires. The monitor in the startup mode may run more than once, but it always runs during the time specified by the stabilization interval value in SMIT. If the monitor returns *within* the stabilization interval, its zero return code indicates that the application had successfully started. If the monitor returns a non-zero code within the stabilization interval, this is interpreted as a failure of the application to start.

  Use this mode for applications in parent resource groups. If you configure dependencies between resource groups in the cluster, the applications in these resource groups are started sequentially as well. To ensure that this process goes smoothly, we recommend configuring several application monitors, and, especially, a monitor that checks the application startup for the application that is included in the parent resource group. This ensures that the application in the parent resource group starts successfully.
- *Long-Running Mode*. In this mode, the monitor periodically checks that the application is running successfully. The checking begins *after* the stabilization interval expires and it is assumed that the application server is started and the cluster has stabilized. The monitor in the long-running mode runs at multiple intervals based on the monitoring interval value that you specify in SMIT.

  Configure a monitor in this mode for any application server. For example, applications included in child and parent resource groups can use this mode of monitoring.
- *Both.* In this mode, the monitor checks for the successful startup of the application server *and* periodically checks that the application is running successfully.

Retry Count and Restart Interval

The restart behavior depends on two parameters, the *retry count* and the *restart interval*, that you configure in SMIT.

- *Retry count*. The retry count specifies how many times HACMP should try restarting before considering the application failed and taking subsequent fallover or notify action.
- *Restart interval*. The restart interval dictates the number of seconds that the restarted application must remain stable before the retry count is reset to zero, thus completing the monitor activity until the next failure occurs.

Note: Do not specify both of these parameters if you are creating an application monitor that will only be used as in a startup monitoring mode.

If the application successfully starts up before the retry count is exhausted, the restart interval comes into play. By resetting the restart count, it prevents unnecessary fallover action that could occur when applications fail several times over an extended time period. For example, a monitored application with a restart count set to three (the default) could fail to restart twice, and then successfully start and run cleanly for a week before failing again. This third failure should be counted as a new failure with three new restart attempts before invoking the fallover policy. The restart interval, set properly, would ensure the correct behavior: it would have reset the count to zero when the application was successfully started and found in a stable state after the earlier failure.

Be careful not to set the restart interval for a too short period of time. If the time period is too short, the count could be reset to zero too soon, before the immediate next failure, and the fallover or notify activity will never occur.

See the instructions for setting the retry count and restart intervals later in this chapter for additional details.

Application Monitoring Prerequisites and Considerations

Keep the following in mind when planning and configuring application monitoring:

- Any application to be monitored must be defined to an application server in an existing cluster resource group.
- If you have configured dependent resource groups, we recommend to configure multiple monitors: for applications included in parent resource groups, and for applications in child resource groups. For example, a monitor for a parent resource group can monitor the successful startup of the application, and a monitor for a child resource group can monitor the process for an application. For more information, see Monitor Modes.
- Multiple monitors can be configured for the same application server. Each monitor can be assigned a unique name in SMIT.
- The monitors that you configure must conform to existing configuration rules. For more information, see Configuring a Process Application Monitor and Configuring a Custom Application Monitor.

- We recommend that you first configure an application server, and then configure the monitor(s) that you can associate with the application server. Before configuring an application monitor, configure all your application servers. Then configure the monitors and associate them with the servers. You can go back at any time and change the association of monitors to servers.
- You can configure no more than 128 monitors per cluster. No limit exists on the number of monitors per application server, as long as the total number of all monitors in the cluster is less than 128.
- When multiple monitors are configured that use different fallover policies, each monitor specifies a failure action of either “notify” or “fallover”. HACMP processes actions in the order in which the monitors indicate an error. For example, if two monitors are configured for an application server and one monitor uses the “notify” method and the other uses the “fallover” method, the following occurs:
    - If a monitor with “fallover” action indicates an error first, HACMP moves the resource group to another node, and the remaining monitor(s) are shut down and restarted on another node. HACMP takes no actions specified in any other monitor.
    - If a monitor with “notify” action indicates an error first, HACMP runs the “notify” method and shuts down that monitor, but any remaining monitors continue to operate as before. You can manually restart the “notify” monitor on that node using the **Suspend/Resume Application Monitoring** SMIT panel.
- If multiple monitors are used, HACMP does not use a particular order for the monitors startup or shutdown. All monitors for an application server are started at the same time. If two monitors are configured with different fallover policies, and they fail at precisely the same time, HACMP does not guarantee it processes methods specified for one monitor before methods for the other.
- The same monitor can be associated with multiple application servers using the **Application Monitor(s)** field in the **Change/Show an Application Server** SMIT panel. You can select a monitor from the picklist.
- If you remove an application monitor, HACMP removes it from the server definition for all application servers that were using the monitor, and indicates which servers are no longer using the monitor.
- If you remove an application server, HACMP removes that server from the definition of all application monitors that were configured to monitor the application. HACMP also sends a message about which monitor will no longer be used for the application. If you remove the last application server in use for any particular monitor, that is, if the monitor will no longer be used for any application, verification issues a warning that the monitor will no longer be used.

Note on the Fallover Option and Resource Group Availability

Be aware that if you select the **fallover** option of application monitoring in the **Customize Resource Recovery** SMIT panel—which could cause a resource group to migrate from its original node—the possibility exists that while the highest priority node is up, the resource group remains inactive. This situation occurs when an **rg_move** event moves a resource group from its highest priority node to a lower priority node, and then you stop the cluster services on the lower priority node with the option to take all the resources offline. Unless you bring the resource group up manually, it remains in an inactive state.

Also, for more information on resource group availability, see the section Selective Fallover for Handling Resource Groups in Appendix B: Resource Group Behavior during Cluster Events.

Steps for Configuring Multiple Application Monitors

To define multiple application monitors for an application:

1. Define one or more application servers. For instructions, see the section Configuring Application Servers in Configuring an HACMP Cluster (Standard).
2. Add the monitors to HACMP. The monitors can be added using the HACMP **Extended Configuration** path in SMIT. For instructions, see the sections Configuring a Process Application Monitor and Configuring a Custom Application Monitor.

Configuring a Process Application Monitor

You can configure multiple application monitors and associate them with one or more application servers. By supporting multiple monitors per application, HACMP can support more complex configurations.

Process application monitoring uses the RSCT subsystem functions to detect the termination of a process and to generate an event. This section describes how to configure process application monitoring, in which you specify one or more processes of a single application to be monitored.

Note: Process monitoring may not be the appropriate solution for all applications. For instance, you cannot monitor a shell script with a process application monitor. If you wish to monitor a shell script, configure a custom monitor. See Configuring a Custom Application Monitor for details on the other method of monitoring applications.

Identifying Correct Process Names

For process monitoring, it is very important that you list the correct process names in the SMIT **Add Process Application Monitor** panel. You should use processes that are listed in response to the **ps -el** command, and not **ps -f**. (This is true for any process that is launched through a `#!<path name>` in the script. For example, this is true for **bsh**, **csh**, etc).

If you are unsure of the correct names, use the following short procedure to identify all the process names for your list.

To identify correct process names:

1. Enter the following command:
   ```
   ps -el | cut -c72-80 | sort > list1
   ```
2. Run the application server.
3. Enter the following command:
   ```
   ps -el | cut -c72-80 | sort > list2
   ```
4. Compare the two lists by entering
   ```
   diff list1 list2 | grep \>
   ```
   The result is a complete and accurate list of possible processes to monitor. You may choose not to include all of them in your process list.

Steps for Configuring a Process Application Monitor

An application must have been defined to an application server before you set up the monitor.

To configure a process application monitor (in any of the three running modes: startup mode, long-running mode or both):

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Resource Configuration** > **HACMP Extended Resources Configuration** > **Configure HACMP Applications > Configure HACMP Application Monitoring > Configure Process Application Monitors > Add Process Application Monitor** and press Enter. A list of previously defined application servers appears.
3. Select the application server to which you want to add a process monitor.
4. In the **Add a Process Application Monitor** panel, fill in the field values as follows:

Monitor Name	Enter the name of the application monitor. Each monitor can have a unique name that does not have to be the same name as the application server name.

Monitor Mode

Select the mode in which the application monitor monitors the application:

- **startup monitoring**. In this mode the application monitor checks that the application server has successfully started *within* the specified stabilization interval. The monitor in this mode may run multiple times, as long as it is being run *within* the stabilization interval that you specify. If the monitor in this mode returns a zero code, this means that the application had started successfully. If a non-zero code is returned, this means that the application did not start within the stabilization interval. Select this mode if you are configuring an application monitor for an application that is included in a parent resource group (in addition to other monitors that you may need for dependent resource groups).
- **long-running monitoring**. In this mode, the application monitor periodically checks that the application server is running. The monitor is run multiple times based on the monitoring interval that you specify. If the monitor returns a zero code, it means that the application is running successfully. A non-zero return code indicates that the application has failed. The checking starts *after* the specified stabilization interval has passed. This mode is the default.
- **both**. In this mode, the application monitor checks that within the stabilization interval the application server has started successfully, *and* periodically monitors that the application server is running after the stabilization interval has passed. If the same monitor is used in the "both" mode, HACMP interprets the return codes differently, according to which type of monitoring is used (see the description of modes).

Processes to Monitor

Specify the process(es) to monitor. You can type more than one process name. Use spaces to separate the names.

Note: To be sure you are using correct process names, use the names as they appear from the **ps -el** command (*not* **ps -f**), as explained in the section Identifying Correct Process Names.

Process Owner

Specify the user ID of the owner of the processes specified above, for example *root*. Note that the process owner must own all processes to be monitored.

Instance Count

Specify how many instances of the application to monitor. The default is **1** instance. The number of instances must exactly match the number of processes to monitor. If you put one instance, and another instance of the application starts, you will receive an application monitor error.

Note: This number *must* be more than **1** if you have specified more than one process to monitor (1 instance for each process).

Stabilization Interval

Specify the time (in seconds). HACMP uses the stabilization period for the monitor in different ways, depending on which monitor mode is selected in this SMIT panel:

- If you select the **startup monitoring** mode, the stabilization interval is the period *within* which HACMP runs the monitor to check that the application has successfully started. When the specified time expires, HACMP terminates the monitoring of the application startup and continues event processing. If the application fails to start within the stabilization interval, the resource group's acquisition fails on the node, and HACMP launches resource group recovery actions to acquire a resource group on another node. The number of seconds you specify should be approximately equal to the period of time it takes for the application to start. This depends on the application you are using.
- If you select the **long-running** mode for the monitor, the stabilization interval is the period during which HACMP waits for the application to stabilize, before beginning to monitor that the application is running successfully. For instance, with a database application, you may wish to delay monitoring until after the start script and initial database search have been completed. You may need to experiment with this value to balance performance with reliability.
- If you select **both** as a monitoring mode, the application monitor uses the stabilization interval to wait for the application to start successfully. It uses the same interval to wait until it starts checking periodically that the application is successfully running on the node.

Note: In most circumstances, this value should *not* be zero.

Restart Count	Specify the number of times to try restarting the application before taking any other actions. The default is **3**. If you are configuring a monitor that is going to be used only in the startup monitoring mode, restart count does not apply, and HACMP ignores values entered in this field. **Note:** Make sure you enter a Restart Method if your Restart Count is any non-zero value.
Restart Interval	Specify the interval (in seconds) that the application must remain stable before resetting the restart count. Do not set this to be shorter than `(Restart Count) x (Stabilization Interval)`. The default is 10% longer than that value. If the restart interval is too short, the restart count will be reset too soon and the desired fallover or notify action may not occur when it should. If you are configuring a monitor that is going to be used only in the startup monitoring mode, restart interval does not apply, and HACMP ignores values entered in this field.
Action on Application Failure	Specify the action to be taken if the application cannot be restarted within the restart count. You can keep the default choice **notify**, which runs an event to inform the cluster of the failure, or select **fallover**, in which case HACMP recovers the resource group containing the failed application on the cluster node with the next highest priority for that resource group. If you are configuring a monitor that is going to be used only in the startup monitoring mode, the action specified in this field does not apply, and HACMP ignores values entered in this field. See Note on the Fallover Option and Resource Group Availability for more information.
Notify Method	(*Optional*) Define a notify method that will run when the application fails. This custom method runs during the restart process and during notify activity. If you are configuring a monitor that is going to be used only in the startup monitoring mode, the method specified in this field does not apply, and HACMP ignores values entered in this field.

Cleanup Method	(*Optional*) Specify an application cleanup script to be called when a failed application is detected, before calling the restart method. The default is the application server stop script defined when the application server was set up (if you have only one application server defined. If you have multiple application servers, enter the stop script in this field that is used for the associated application server). If you are configuring a monitor that is going to be used only in the startup monitoring mode, the method specified in this field does not apply, and HACMP ignores values entered in this field. **Note**: With application monitoring, since the application is already stopped when this script is called, the server stop script may fail.
Restart Method	(Required if **Restart Count** is not zero.) The default restart method is the application server start script defined previously, if only one application server was set up. This field is empty if multiple servers are defined. You can specify a different method here if desired. If you are configuring a monitor that is going to be used only in the startup monitoring mode, the method specified in this field does not apply, and HACMP ignores values entered in this field.

5. Press Enter.

SMIT checks the values for consistency and enters them into the HACMP Configuration Database. When the resource group is brought online, the application monitor in the long-running mode starts (if it is defined). Note that the application monitor in the startup monitoring mode starts before the resource group is brought online.

When you synchronize the cluster, verification ensures that all methods you have specified exist and are executable on all nodes.

Configuring a Custom Application Monitor

You can configure multiple application monitors and associate them with one or more application servers. By supporting multiple monitors per application, HACMP can support more complex configurations. For more information, see the section Steps for Configuring Multiple Application Monitors.

Custom application monitoring allows you to write a monitor method to test for conditions other than process termination. For example, if an application sometimes becomes unresponsive while still running, a custom monitor method could test the application at defined intervals and report when the application's response is too slow. Also, some applications (`shell` scripts, for example) cannot be registered with RSCT, so process monitoring cannot be configured for them. A custom application monitor method can monitor these types of applications.

For instructions on defining a process application monitor, which requires no custom monitor method, refer to Application Monitoring Prerequisites and Considerations.

Notes on Defining a Monitor Method

Unlike process monitoring, custom application monitoring requires you to provide a script to test the health of the application. You must also decide on a suitable polling interval.

When devising your custom monitor method, keep the following points in mind:

- The monitor method must be an executable program (it can be a `shell` script) that tests the application and exits, returning an integer value that indicates the application's status. The return value must be zero if the application is healthy, and must be a non zero value if the application has failed.
- HACMP cannot pass arguments to the monitor method.
- The method can log messages to **/tmp/clappmond.*application monitor name*.monitor.log** by simply printing messages to the standard output (**stdout**) file. The monitor log file is overwritten each time the application monitor runs.
- Since the monitor method is set up to terminate if it does not return within the specified polling interval, do not make the method overly complicated.

Steps for Configuring a Custom Application Monitor

To set up a custom application monitoring method:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Resource Configuration** > **HACMP Extended Resources Configuration** > **Configure HACMP Applications > Configure HACMP Application Monitoring > Configure Custom Application Monitors > Add a Custom Application Monitor** and press Enter.

   A list of defined application servers appears.
3. Select the application server for which you want to add a monitoring method.
4. In the **Add Custom Application Monitor** panel, fill in field values as follows. Note that the **Monitor Method** and **Monitor Interval** fields require you to supply your own scripts and specify your own preference for the polling interval:

Application Server Name	Select the application server from the picklist.

Monitor Mode

Select the mode in which the application monitor will monitor the application:

- **startup monitoring**. In this mode the application monitor checks that the application server has successfully started *within* the specified stabilization period. If you are configuring a monitor for an application that is included in a parent resource group, select this mode (in addition to other monitors that you may need for dependent resource groups).
- **long-running monitoring**. In this mode, the application monitor periodically checks that the application server is running. The checking starts *after* the specified stabilization interval has passed. This is the default.
- **both**. In this mode, the application monitor checks that within the stabilization interval the application server has started successfully, and periodically monitors that the application server is running after the stabilization interval have passed.

Monitor Method

Enter a script or executable for custom monitoring of the health of the specified application. Do not leave this field blank.

Note that the method must return a zero value if the application is healthy and a non-zero value if a problem is detected.

You can have the monitor log messages to the log file **/tmp/clappmond.*application monitor name*.monitor.log** by having it print messages to the **stdout** file. The messages are automatically redirected to the monitor log.

Monitor Interval

Enter the polling interval (in seconds) for checking the health of the application. If the monitor does not respond within this interval, it is considered hung.

Hung Monitor Signal

The signal the system should send to stop the **Monitor Method** script if it does not return within the time specified for the **Monitor Interval**. The default is SIGKILL(9).

Stabilization Interval

Specify the time (in seconds). HACMP uses the stabilization period for the monitor in different ways, depending on which monitor mode is selected in this SMIT panel:

- If you select the **startup monitoring** mode, the stabilization interval is the period within which HACMP monitors that the application has successfully started. When the specified time expires, HACMP terminates the monitoring of the application startup, and continues event processing. If the application fails to start within the stabilization interval, the resource group's acquisition fails on the node, and HACMP launches resource group recovery actions to acquire a resource group on another node. The number of seconds you specify should be approximately equal to the period of time it takes for the application to start. This depends on the application you are using.
- If you select the **long-running** mode for the monitor, the stabilization interval is the period during which HACMP waits for the application to stabilize, before beginning to monitor that the application is running successfully. For instance, with a database application, you may wish to delay monitoring until after the start script and initial database search have been completed. You may need to experiment with this value to balance performance with reliability.
- If you select **both** as a monitoring mode, the application monitor uses the stabilization interval to wait for the application to start successfully. It uses the same interval to wait until it starts checking periodically that the application is successfully running on the node.

Note: In most circumstances, this value should *not* be zero.

Restart Count

Specify the number of times to try restarting the application before taking any other actions. The default is **3**.

Restart Interval

Specify the interval (in seconds) that the application must remain stable before resetting the restart count. Do not set this to be shorter than (Restart Count) x (Stabilization Interval + Monitor Interval). The default is 10% longer than that value. If the restart interval is too short, the restart count will be reset too soon and the desired failure response action may not occur when it should.

Action on Application Failure	Specify the action to be taken if the application cannot be restarted within the restart count. You can keep the default choice **notify**, which runs an event to inform the cluster of the failure, or select **fallover**, in which case the resource group containing the failed application moves over to the cluster node with the next highest priority for that resource group.
Notify Method	(*Optional*.) The full pathname of a user defined method to perform notification when a monitored application fails. This method will execute each time an application is restarted, fails completely, or falls over to the next node in the cluster. Configuring this method is strongly recommended.
Cleanup Method	(*Optional*) Specify an application cleanup script to be invoked when a failed application is detected, before invoking the restart method. The default is the application server stop script defined when the application server was set up. **Note**: With application monitoring, since the application may be already stopped when this script is called, the server stop script may fail. For more information on stop scripts, see Appendix B: Applications and HACMP in the *Planning Guide*.
Restart Method	(Required if **Restart Count** is not zero.) The default restart method is the application server start script defined previously, when the application server was set up. You can specify a different method here if desired.

5. Press Enter.

SMIT checks the values for consistency and enters them into the HACMP Configuration Database. When the resource group comes online, the application monitor in the long-running mode starts. (The application startup monitor starts before the resource group is brought online).

When you synchronize the cluster, verification ensures that all methods you have specified exist and are executable on all nodes.

Suspending, Changing, and Removing Application Monitors

You can temporarily suspend an application monitor in order to perform cluster maintenance. You should not change the application monitor configuration while it is in a suspended state.

If you have multiple application monitors configured, and choose to temporarily suspend an application monitor, all monitors configured for a specified server are suspended.

For instructions on temporarily suspending application monitoring, changing the configuration, or permanently deleting a monitor, see the section Changing or Removing Application Monitors in Managing the Cluster Resources.

Configuring Tape Drives as HACMP Resources

The HACMP SMIT panels enable the following actions for configuring tape drives:

- Add tape drives as HACMP resources
  - Specify synchronous or asynchronous tape operations
  - Specify appropriate error recovery procedures
- Change or show tape drive resources
- Remove tape drive resources
- Add tape drives to HACMP resource groups
- Remove tape drives from HACMP resource groups.

Adding a Tape Resource

To add a tape drive as a cluster resource:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Resource Configuration > HACMP Extended Resources Configuration > Configure HACMP Tape Resources** > **Add a Tape Resource** and press Enter.
3. Enter the field values as follows:

Tape Resource Name	The symbolic name for the tape resource. This is a required field, and must be unique within the cluster. The name can have up to 32 alphanumeric characters and underscores.
Description	Description of the tape resource.
Tape Device Name	The name of the special file for the tape drive, for example, /**dev/rmt0**. This is a required field.
Start Script	The script to execute on tape resource startup. The default is no user-provided script execution.
Start Processing Synchronous?	If **yes,** then tape start processing is synchronous. If **no**, it is asynchronous. The default is synchronous operation.
Stop Script	The script to execute on tape resource shutdown. The default is no user-provided script execution.
Stop Processing Synchronous?	If **yes**, then tape stop processing is synchronous. If **no**, it is asynchronous. The default is synchronous operation.

Sample scripts are available in the **/usr/es/sbin/cluster/samples/tape** directory. The sample scripts rewind the tape drive explicitly. See Appendix A: Script Utilities in the *Troubleshooting Guide* for the syntax.

Change or Show a Tape Resource

To change or show the current configuration of a tape drive resource, see Reconfiguring Tape Drive Resources in Chapter 14: Managing the Cluster Resources.

Adding a Tape Resource to a Resource Group

To add a tape drive resource to a resource group:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Resource Configuration > HACMP Extended Resource Group Configuration** > **Change/Show Resources and Attributes for a Resource Group** and press Enter.

   SMIT displays the list of resource groups.
3. Select the resource group to which you want to add the tape resource.

   SMIT displays the **Change/Show all Resources/Attributes for a <selected type of> Resource Group** panel.
4. Enter the field value for **Tape Resource**.

   Type in the resource name or press F4 to display a picklist of defined tape resources. Select the desired resource. If there are no tape resources defined, SMIT displays an error message.

Verifying and Synchronizing Tape Drive Configuration

After adding a resource to a resource group, verify that the configuration is correct and then synchronize shared tape resources to all nodes in the cluster.

Verification ensures the following:

- Validity of the specified tape special file (is it a tape drive?)
- Accessibility of the tape drive (does a device on the specified SCSI LUN exist?)
- Consistency of the configuration (does the device have the same LUN on the nodes sharing the tape drive?)
- Validity of the user defined start and stop scripts (do the scripts exist and are they executable?)

Dynamic Reconfiguration of Tape Resources

When a tape drive is added to a resource group, or when a new resource group is created with tape resources, DARE will reserve the tape and invoke the user-provided tape start script.

When a tape drive is removed from a resource group, or when a resource group with tape resources is removed, DARE invokes the user-provided tape stop script and releases the tape drive.

Configuring AIX 5L Fast Connect

The AIX 5L Fast Connect for Windows application is integrated with HACMP, so you can configure it, via the SMIT interface, as a highly available resource in resource groups. This application does not need to be associated with application servers or special scripts.

Refer to your planning worksheets as you prepare to configure the application as a resource.

AIX 5L Fast Connect allows client PCs running Windows, DOS, and OS/2 operating systems to request files and print services from an AIX 5L server. Fast Connect supports the transport protocol NetBIOS over TCP/IP.

Prerequisites

Before you can configure Fast Connect resources in HACMP, make sure these steps have been taken:

- Install the Fast Connect Server on all nodes in the cluster.
- AIX 5L print queue names match for all nodes in the cluster if Fast Connect printshares are to be highly available.
- For non-concurrent resource groups, assign the *same* netBIOS names to each node. when the Fast Connect Server is configured. This action will minimize the steps needed for the client to connect to the server after fallover.
- For concurrently configured resource groups, assign *different* netBIOS names across nodes.
- Configure on the Fast Connect Server those files and directories on the AIX 5L machine that you want shared.

Configuration Notes for Fast Connect

When configuring Fast Connect as a cluster resource in HACMP, keep the following points in mind:

- When starting cluster services, the Fast Connect server should be stopped on all nodes, so that HACMP can take over the starting and stopping of Fast Connect resources properly.
- In concurrent configurations, the Fast Connect server should have a second, non-concurrent, resource group defined that does not have Fast Connect on it. Having a second resource group configured in a concurrent cluster keeps the AIX 5L filesystems used by Fast Connect cross-mountable and highly available in the event of a node failure.
- Fast Connect cannot be configured in a mutual takeover configuration. Make sure there are no nodes participating in more than one Fast Connect resource groups at the same time.

For instructions on using SMIT to configure Fast Connect services as resources, see the section Adding Resources and Attributes to Resource Groups Using the Extended Path in Chapter 5: Configuring HACMP Resource Groups (Extended).

Verification of Fast Connect

After completing your resource configuration, you synchronize cluster resources. During this process, if Fast Connect resources are configured in HACMP, the verification ensures that:

- The Fast Connect server application exists on all participating nodes in a resource group.
- The Fast Connect fileshares are in filesystems that have been defined as resources on all nodes in the resource group.
- The Fast Connect resources are not configured in a mutual takeover form; that is, there are no nodes participating in more than one Fast Connect resource group.

Configuring Highly Available Communication Links

HACMP can provide high availability for three types of communication links:

- SNA configured over a LAN interface
- SNA over X.25
- Pure X.25.

Highly available SNA links can use either LAN interfaces (as in previous releases of HACMP) or X.25 links.

LAN interfaces are Ethernet, Token Ring, and FDDI interfaces; these interfaces are configured as part of the HACMP cluster topology.

X.25 interfaces are usually, though not always, used for WAN connections. They are used as a means of connecting dissimilar machines, from mainframes to dumb terminals. Because of the way X.25 networks are used, these interfaces are treated as a different class of devices that are *not* included in the cluster topology and *not* controlled by the standard HACMP topology management methods. This means that heartbeats are not used to monitor X.25 interface status, and you do not define X.25-specific networks in HACMP. Instead, an HACMP daemon, **clcommlinkd,** takes the place of heartbeats. This daemon monitors the output of the **x25status** command to make sure the link is still connected to the network.

Basic Steps for Configuring Highly Available Communication Links

Making a communication link highly available in HACMP involves these general steps:

1. Define the communication interfaces and links in AIX 5L.
2. Define the interfaces and links in HACMP.
3. Add the defined communication links as resources in HACMP resource groups.

These steps will be explained further in the sections covering each of the three communication link types.

Configuring SNA-Over-LAN Communication Links

Using the SMIT interface, you can configure an SNA link in AIX 5L and HACMP and add the link to a resource group.

Supported Software Versions

To configure highly available SNA links, you must have Communication Server for AIX 5L (CS/AIX) version 6.1 or higher.

Creating a Highly Available SNA-over-LAN Communication Link

To configure a highly available SNA-over-LAN communication link, you configure the link first in AIX 5L, then in HACMP, and finally you add the link to a cluster resource group; these procedures are described in the following sections.

Note that the AIX 5L configuration steps must be performed on each node; the HACMP steps can be performed on a single node and then the information will be copied to all cluster nodes during synchronization.

Configuring the SNA Link in AIX 5L

To define an SNA link in AIX 5L:

1. Enter `smit hacmp`
2. In SMIT, select **System Management (C-SPOC) > HACMP Communication Interface Management > Configure Communication Interfaces/Devices to the Operating System on a Node** and press Enter.

3. Select a node from the list.
4. Select a network interface type **SNA Communication Links** from the list.

   If you have the Communication Server for AIX 5L (CS/AIX) version 6.1 or higher installed, this brings you to the main AIX 5L SMIT menu for SNA system configuration. Press F1 for help on entries required for configuring these links. Note that a valid SNA configuration must exist before an SNA link can be configured to HACMP and made highly available.

Note: It is strongly recommended that you test the link outside of HACMP before adding it to HACMP.

Configuring the SNA Link in HACMP

To configure a link in HACMP:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Resource Configuration > HACMP Extended Resources Configuration > Configure HACMP Communication Adapters and Links > Configure Highly Available Communication Links > Add Highly Available Communication Link > Add Highly Available SNA-over-LAN link** and press Enter.

   Enter field values as follows:

Name	Enter the name by which you want the link to be known to HACMP throughout the cluster. This name must be unique among all highly available communication links, regardless of type, within the cluster. It can include alphanumeric characters and underscores, but cannot have a leading numeric. Maximum size is 32 characters.
DLC Name	Specify the SNA Data Link Control (DLC) profile to be made highly available.
Port(s)	Enter ASCII text strings for the names of any SNA ports to be started automatically.
Link Station(s)	Enter ASCII text strings for the names of the SNA link stations.
Application Service File	Enter the name of the file that this link should use to perform customized operations when this link is started or stopped. For more information on how to write an appropriate script, see Notes on Application Service Scripts for Communication Links.

3. After entering all values, press Enter to add this information to the HACMP Configuration Database.

Important Notes

- It is possible to configure a DLC in HACMP when no connection actually exists. There is no guarantee that the port(s) or link station(s) are actually there. In other words, HACMP does not monitor the SNA DLCs, ports, and links directly for actual status—it monitors only the service interface over which SNA is running. Therefore, the SNA-over-LAN connection is only as highly available as the interface.
- The DLC follows the service label when the resource group falls over, so make sure you include a service label in the resource group that holds the SNA-over-LAN link.
- All SNA resources (DLC, ports, link stations) must be properly configured in AIX 5L when HACMP cluster services start up, in order for HACMP to start the SNA link and treat it as a highly available resource. If an SNA link is already running when HACMP starts, HACMP stops and restarts it.
- If multiple SNA links are defined in the operating system, make sure *all* of them are defined to HACMP. If not, HACMP will still stop them all at startup, but the links outside of HACMP will not be restarted.

Adding the SNA Link to a Resource Group

You may or may not have resource groups defined at this point. The process for creating resource groups and adding resources to them is covered in Chapter 5: Configuring HACMP Resource Groups (Extended).

Note: When a resource group has a list of Service IP labels and Highly Available Communication Links with configured SNA resources, the first Service IP label in the list of Service IP labels defined in the resource group will be used to configure SNA.

To complete the configuration of highly available SNA-over-LAN communication links, you must add them to a resource group.

To add the SNA Link to a resource group:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Resource Configuration > HACMP Extended Resource Group Configuration > Change/Show Resources and Attributes for a Resource Group** and press Enter. SMIT displays a list of resource groups.
3. Select a resource group.
4. Specify the links in the **Communication Links** field for the resource group.
5. Press Enter.

Changing or Removing an SNA Communication Link

To change or remove a highly available SNA-over-LAN communication link, see the relevant SMIT options on the **Extended Configuration >Extended Resource Configuration** menu.

SNA Communication Links as Highly Available Resources

HACMP protects SNA-over-LAN connections during interface and node failures. This section describes the HACMP actions that take place for each of these failures.

Network Interface Failure

When a service interface over which SNA is running fails, HACMP will take the following actions:

1. Stops the DLC, which stops the link stations and the ports
2. Modifies the DLC to use an available standby interface
3. Restarts the ports, the link stations, and the DLC.

Depending on the number of DLC ports and link stations, this process may take several minutes. The DLC and its associated ports and link stations may be unavailable during the time it takes to recover from an interface failure. Clients or applications connected before the failure may have to reconnect.

If no standby interface is available, the event is promoted to a local network failure event, causing the affected resource group to fallover to another node, if another node using the same network is available.

Node Failure

Once any communication link type is configured as a resource, it is treated like other cluster resources in the event of a node failure. When a node fails, the resource group is taken over in the normal fashion, and the link is restarted on the takeover node. Any identified resources of that link, such as link stations and ports, are restarted on the takeover node.

Network Failure

Network failures are handled as they are in a non-SNA environment. When a network failure occurs, HACMP detects an IP **network_down** event and logs an error message in the **/tmp/hacmp.out** file. Even though the SNA connection is independent of the IP network, it is assumed that an IP network failure event indicates that the SNA link is also down. The local **network_down** event causes the resource group containing the SNA link to fall over to another node, if another node using the same network is available.

Verification of SNA Communication Links

Verification ensures that:

- The specified DLCs, ports, and links exist and are correctly associated with each other
- The specified application service file exists and is readable and executable.

There is no checking for invalid SNA configuration information; it is assumed that the system administrator has properly configured and tested SNA.

Note: Verification will fail if the SNA server is not running when verification is run. If SNA is stopped on any node in the resource group at the time of verification, HACMP reports an error, even if the SNA DLC is properly configured.

Configuring X.25 Communication Links

An X.25 communication link can be made highly available when included as a resource in an HACMP resource group. It is then monitored by the **clcommlinkd** daemon and appropriate fallover actions are taken when an X.25 communication link fails.

Supported Adapters and Software Versions

The following adapters are supported for highly available X.25 communication links:

- IBM 2-Port Multiprotocol Adapter (DPMP)
- IBM Artic960Hx PCI Adapter (Artic)

To configure highly available X.25 links, you must have AIXLink/X.25 version 2 or higher.

Creating an X.25 Communication Link

These steps describe how to configure a highly available X.25 communication link. You must first configure the X.25 adapter in AIX 5L, then configure the adapter and the link in HACMP, and finally, add the link to an HACMP resource group.

Note that the AIX 5L configuration steps must be performed on each node; the HACMP steps can be performed on a single node and then the information will be copied to all cluster nodes during synchronization.

WARNING: HACMP should handle the starting of your X.25 links. If HACMP attempts to start the link and finds that it already exists, link startup fails, because all ports must have unique names and addresses. For this reason, make sure the X.25 port is not defined to AIX 5L when the cluster starts.

It is highly recommended that you configure the adapters and drivers and test the X.25 link outside of HACMP before adding it to HACMP. However, if you do this, you must make sure to delete the X.25 port from AIX 5L before starting HACMP, so that HACMP can properly handle the link startup.

Configuring the X.25 Adapter in AIX 5L

To configure an X.25 communication link you first configure the adapter in AIX 5L as follows:

1. Enter `smit hacmp`
2. In SMIT, select **System Management (C-SPOC) > HACMP Communication Interface Management > Configure Communication Interfaces/Devices to the Operating System on a Node** and press Enter.
3. Select a node from the list.
4. Select a network interface type **X.25 Communication Interfaces** from the list.

   If you have the Communication Server for AIX 5L (CS/AIX) version 6.1 or higher installed, this brings you to an AIX 5L menu listing adapter types.
5. Select an adapter and fill in the **Adapter**, **Services**, and **User Applications** fields.

Configuring the X.25 Adapter in HACMP

After you define the adapter in AIX 5L, you must configure the adapter and the link in HACMP as follows:

1. In SMIT, enter `smit hacmp`

2. In SMIT, select **Extended Configuration > Extended Resource Configuration > HACMP Extended Resources Configuration** > **Configure HACMP Communication Adapters and Links > Configure Communication Adapters for HACMP > Add Communication Adapter** and press Enter.
3. Enter field values as follows:

Name	Enter the name by which the adapter will be known throughout the cluster. This name must be unique among all communication adapters within the cluster. It can include alphanumeric characters and underscores, but cannot have a leading numeric.The maximum size is 32 characters.
Node	Press F4 for a picklist of node names, and specify the node on which the adapter is installed and configured. This node must already be part of the cluster.
Device	Enter the device-file name of the driver used by this adapter/port. For an Artic adapter, this name will be of the format **twd#;** for a DPMP adapter, it will be **hdlc#.** The driver must exist on the specified node.
Port Number	For cards that use a twd driver (such as Artic), specify the adapter port number. For cards that use an hdlc driver (such as DPMP), this field is not required and will be ignored.
Multiple Links Allowed	Set this field to **true** if you want to be able to list this adapter as available to more than one highly available communication link. Leave it set to **false** if you know that only one communication link should be able to use this adapter.
Allowable Links	(*Optional*) Set this field to one or more names of highly available communication links (separated by blanks) if you know that this adapter should used only by certain links. Otherwise, leave it blank. Note that this field cannot be defined until after the communication link has been added to an HACMP resource group. A warning is issued if you enter information here before the link is defined in a resource group.

4. Press Enter to add this information to the HACMP Configuration Database.

Configuring the X.25 Link in HACMP

After you configure the X.25 adapter in HACMP, you must configure the X.25 link in HACMP as follows:

1. Enter `smit hacmp`

2. In SMIT, select **Extended Configuration > Extended Resource Configuration > HACMP Extended Resources Configuration > Configure HACMP Communication Adapters and Links > Configure Highly Available Communication Links > Add Highly Available Communication Link > Add Highly Available X.25 Link** and press Enter.
3. Enter field values as follows:

Name	Enter the name by which you want the link to be known to HACMP throughout the cluster. This name must be unique among all highly available communication links, regardless of type, within the cluster. It can include alphanumeric characters and underscores, but cannot have a leading numeric.The maximum size is 32 characters.
Port	Enter the X.25 port designation you wish to use for this link, for example `sx25a0`. The port name must be unique across the cluster.This name must begin with "`sx25a`" but the final numeric character is your choice. The port name can be up to 8 characters long; therefore, the final numeric can contain up to three digits.
Address/NUA	Enter the X.25 address (local NUA) that will be used by this link.
Network ID	Enter the X.25 network ID. The default value is 5, which will be used if this field is left blank.
Country Code	Enter the X.25 country code. The system default will be used if this field is left blank.
Adapter Name(s)	Press F4 and select from the picklist the communication adapters that you want this link to be able to use. Note that these are the HACMP names, not the device names.
Application Service File	Enter the name of the file that this link should use to perform customized operations when this link is started and/or stopped. For more information on how to write an appropriate script, see Notes on Application Service Scripts for Communication Links.

4. Press Enter to add this information to the HACMP Configuration Database.

Adding the X.25 Link to a Resource Group

You may not have any resource groups defined at this point. The process for creating resource groups and adding resources to them is covered in Chapter 5: Configuring HACMP Resource Groups (Extended).

To complete the configuration of highly available X.25 communication links, you add them to a resource group.

1. Enter `smit hacmp`

2. In SMIT, select **Extended Configuration > Extended Resource Configuration** > **HACMP Extended Resource Group Configuration > Change/Show Resources and Attributes for a Resource Group** and press Enter.
3. Select the resource group.
4. Specify the links in the **Communication Links** field for the resource group.

Changing or Removing an X.25 Communication Link

For information on changing or removing a highly available X.25 communication link, see Reconfiguring Communication Links in Chapter 14: Managing the Cluster Resources.

X.25 Communication Links as Highly Available Resources

X.25 connections are protected during network interface and node failures. This section describes the HACMP actions that take place for each of these failures.

Network Adapter Failure

The X.25 link is monitored by an HACMP daemon, **clcommlinkd**. The daemon monitors whether the port is connected to the network, by checking the output of **x25status**. It does not check for an actual connection between nodes. If network connectivity is lost, the link falls over to another adapter port on the same node, or—if another port is not available on the node—the affected resource group falls over to the node with the next highest priority.

Node Failure

Once any communication link type is configured as a resource, it is treated like other cluster resources in the event of a node failure. When a node fails, the resource group is taken over in the normal fashion, and the link is restarted on the takeover node. Any identified resources of that link, such as link stations and ports, are restarted on the takeover node.

Network Failure

Network failures are handled as they would in a non-X.25 environment. When a network failure occurs, HACMP detects an IP network down and logs an error message in the **/tmp/hacmp.out** file. The local network failure event, `network_down <node name> <network name>`, causes the affected resource group to fall over to another node.

Verification of X.25 Communication Links

The HACMP cluster verification process ensures the following:

- The specified adapters exist on the specified nodes.
- There is at least one adapter defined for every node defined in the resource group (since all nodes must be able to acquire the link).
- An adapter has not been specified for more than one link (this check occurs if the **Multiple Links Allowed** field is set to **false).**
- The application service file exists and is readable and executable.

There is no checking for invalid X.25 configuration information; it is assumed that the system administrator has properly configured X.25.

Configuring SNA-Over-X.25 Communication Links

Configuring communication links for SNA running over X.25 involves a combination of the steps for SNA-over-LAN and pure X.25.

Supported Adapters and Software Versions

The following adapters are supported for highly available X.25 communication links:

- IBM Dual Port Multiprotocol Adapter (DPMP)
- Artic960hx 4-Port Adapter (Artic).

To configure highly available SNA-over-X.25 links, you must have the following software:

- Communication Server for AIX 5L (CS/AIX) version 6.1 or higher
- AIXLink/X.25 version 2 or higher.

Creating an SNA-over-X.25 Communication Link

These steps describe how to configure a highly available SNA-over-X.25 communication link.

Note that the AIX 5L configuration steps must be performed on each node; the HACMP steps can be performed on a single node and then the information will be copied to all cluster nodes during synchronization.

To create a highly available SNA-over-X.25 link, follow these steps:

1. Configure the SNA link in AIX 5L only (*not* in HACMP) on each node. For instructions, see Configuring the SNA Link in AIX 5L.
2. Configure the X.25 adapter in AIX 5L on each node. For instructions, see Configuring the X.25 Adapter in AIX 5L.
3. Configure the X.25 adapter in HACMP. For instructions, see Configuring the X.25 Adapter in HACMP.

   WARNING: HACMP should manage the startup of your X.25 links. If HACMP attempts to start the link and finds that it already exists, link startup fails, because all ports must have unique names and addresses. For this reason, make sure the X.25 port is not defined to AIX 5L when the cluster starts. Only the X.25 devices and drivers should be defined at cluster startup.

   It is highly recommended that you configure the adapters and drivers and test the X.25 link outside of HACMP before adding it to HACMP. However, if you do this, you must make sure to delete the X.25 port from AIX 5L before starting HACMP, so that HACMP can properly handle the link startup.

   In contrast, all SNA resources (DLC, ports, link stations) *must* be properly configured in AIX 5L when HACMP cluster services start up, in order for HACMP to start the SNA link and treat it as a highly available resource. If an SNA link is already running when HACMP starts, HACMP stops and restarts it. Also see the Important Notes in the SNA-over-LAN configuration section.

4. Configure the SNA-over-X.25 link in HACMP. From the main HACMP menu, In SMIT, select **Extended Configuration > Extended Resource Configuration > HACMP Extended Resources Configuration** > **Configure HACMP Communication Adapters and Links > Configure Highly Available Communication Links > Add Highly Available Communication Links > Add Highly Available SNA-over-X.25 Link** and press Enter.
5. Enter field values as follows:

Name	Enter the name by which you want the link to be known to HACMP throughout the cluster. This name must be unique among all highly available communication links, regardless of type, within the cluster. It can include alphanumeric characters and underscores, but cannot have a leading numeric. The maximum length is 30 characters.
X.25 Port	Enter the X.25 port designation you wish to use for this link, for example: sx25a0. The port name must be unique across the cluster.This name must begin with “sx25a” but the final numeric character is your choice. The port name can be up to eight characters long; therefore the final numeric can contain up to three digits.
X.25 Address/NUA	Enter the X.25 Address/NUA that will be used by this link.
X.25 Network ID	Enter the X.25 Network ID. The default value is 5, which will be used if this field is left blank.
X.25 Country Code	Enter the X.25 Country Code. The system default will be used if this field is left blank.
X.25 Adapter Name(s)	Enter the HACMP names (not the device names) of the communication adapters that you want this link to be able to use.
SNA DLC	Identify the SNA DLC profile to be made highly available. Pick F4 to see a list of the DLC names.
SNA Port(s)	Enter ASCII text strings for the names of any SNA ports to be started automatically.
SNA Link Station(s)	Enter ASCII text strings for the names of the SNA link stations.
Application Service File	Enter the name of the file that this link should use to perform customized operations when this link is started and/or stopped. For more information on how to write an appropriate script, see Notes on Application Service Scripts for Communication Links.

6. Press Enter to add this information to the HACMP Configuration Database.

Adding the SNA-Over-X.25 Link to a Resource Group

You may or may not have resource groups defined at this point. The process for creating resource groups and adding resources to them is described in Chapter 5: Configuring HACMP Resource Groups (Extended).

To complete the configuration of a highly available SNA-over-X.25 communication link, you add it to a resource group.

1. In SMIT, select **Extended Resource Configuration** > **HACMP Extended Resource Group Configuration > Change/Show Resources and Attributes for a Resource Group** and press Enter. SMIT displays a list of resource groups.
2. Select a resource group.
3. Specify the link in the **Communication Links** field.

Changing or Removing an SNA-Over-X.25 Communication Link

To change or remove a highly available SNA-over-X.25 communication link, see Reconfiguring Communication Links in Chapter 14: Managing the Cluster Resources.

SNA-Over-X.25 Communication Links as Highly Available Resources

SNA-over-X.25 connections are protected during adapter and node failures. This section describes the HACMP actions that take place for each of these failures.

Adapter Failure

The X.25 link is monitored by an HACMP daemon, **clcommlinkd**. The daemon monitors whether the port is connected to the network, by checking the output of **x25status**. It does not check for an actual connection between nodes. When an X.25 link over which SNA is running fails, HACMP causes the link to fall over to another adapter port on the same node, or—if another port is not available on the node—the affected resource group falls over to the node with the next highest priority.

Node Failure

Once any communication link type is configured as a resource, it is treated like other cluster resources in the event of a node failure. When a node fails, the resource group is taken over in the normal fashion, and the link is restarted on the takeover node. Any identified resources of that link, such as link stations and ports, are restarted on the takeover node.

Network Failure

Network failures are handled as they are in a non-SNA/X.25 environment. When a network failure occurs, HACMP detects an IP network down and logs an error message in the **/tmp/hacmp.out** file. The local network failure event, `network_down <node name> <network name>`, causes the resource group containing the link to fall over to another node.

Verification of SNA-Over-X.25 Communication Links

Verification ensures the following:

- The specified SNA DLCs, ports, and links exist and are correctly associated with each other.
- The specified X.25 adapters exist on the specified nodes.

- There is at least one adapter for every node defined in the resource group (since all nodes must be able to acquire the link).
- (If the **Multiple Links Allowed** field is set to **false**) An adapter has not been specified for more than one link.
- The application service file exists and is readable and executable.

There is no checking for invalid SNA or X.25 configuration information; it is assumed that the system administrator has properly configured SNA and X.25.

Note: Verification will fail if the SNA server is not running when verification is run. If SNA is stopped on any node in the resource group at the time of verification, HACMP reports an error ("The DLC <name> is not defined to CS/AIX on node <name>"), even if the SNA DLC is properly configured.

Notes on Application Service Scripts for Communication Links

When you define a communication link in HACMP, you specify the name of an application service file. The application service file may contain a script to perform customized operations when the link is started or stopped by HACMP. This script should contain a START line followed by startup instructions and a STOP line followed by stop instructions.

HACMP passes a single parameter in that identifies the type of link (SNA or X.25) to start or stop. For SNA-over-LAN and pure X.25 links, this parameter does not need to appear in your script. However, for SNA-over-X.25, you must specify "X25" and "SNA" so that the appropriate start and stop scripts run for both link types.

Here is an example application service file for an SNA-over-X.25 link:

```
START
if [[ "$1" = "X25" ]]; then
/usr/local/scripts/my_x25link_app_start_script.sh &
elif [[ "$1" = "SNA" ]]; then
/usr/local/scripts/my_snalink_app_start_script.sh &
fi
STOP
if [[ "$1" = "X25" ]]; then
/usr/local/scripts/my_x25link_app_stop_script.sh &
elif [[ "$1" = "SNA" ]]; then
/usr/local/scripts/my_snalink_app_stop_script.sh &
fi
```

When the application is to be started, all lines between the START and STOP tags are run. When the application is to be stopped, all lines after the STOP tag are run.

For an SNA-over-LAN or an X.25 link, the file need only contain the individual start and stop instructions, as in this example (for SNA-over-LAN):

```
START
/usr/local/scripts/my_snalink_app_start_script.sh &
STOP
/usr/local/scripts/my_snalink_app_stop_script.sh &
```

Customizing Resource Recovery

HACMP monitors system resources and initiates recovery when a failure is detected. Recovery involves moving a set of resources (grouped together in a resource group) to another node. HACMP uses *selective fallover* function when it can. Selective fallover enables HACMP to recover only those resource groups that are affected by the failure of a specific resource.

HACMP uses selective fallover in the following cases:

- Loss of a volume group
- Local network failure
- Resource group acquisition failure
- Application failure.

If you have configured resource groups with sites and an HACMP/XD disaster recovery solution, these are *replicated resources*. HACMP tracks and handles recovery of resource groups on both the primary and the secondary (backup) site. HACMP tries to recover the failure of a secondary instance as well as the primary instance.

You can customize recovery for two types of resources where HACMP uses selective fallover:

- *Service IP labels*. By default, for a local network failure, HACMP responds by scanning the configuration for any service labels on that network and moving only the resource group containing the failed service IP label to another available node.

  Note: You cannot customize recovery for service IP labels for the secondary instance of a replicated resource group.

- *Volume groups*. For volume groups where the recovery is triggered by an AIX 5L error notification from the LVM (caused by a loss of quorum for a volume group), HACMP moves the resource group to a takeover node.

  Note: Customizing volume group recovery (disabling selective fallover) in a cluster with this type of resource in a replicated resource group applies to both the primary and the secondary instances of the resource group.

However, selective fallover may not be the behavior you want when one of these resources fails. After upgrading from a previous release, if you have custom pre- and post-events to handle these situations, these may act in unexpected ways when combined with the selective fallover behavior. HACMP includes the **Customize Resource Recovery** option for changing the behavior of the selective fallover action for these resources. You can select to have the fallover occur, or to simply receive a notification.

Take the following steps to customize resource recovery for service label and volume group resources (especially if you have your own custom pre- and post-event scripts):

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Resource Configuration > HACMP Extended Resources Configuration > Customize Resource Recovery** and press Enter.
3. Select the resource to customize from the list.

4. Enter field values as follows:

Name	The resource you selected.
Action on Resource failure	Select either **fallover** or **notify**. **Fallover** is the default. • **Fallover** initiates an **rg_move** event to move the affected resource group to another node. • **Notify** causes a **server_down** event that calls out the specific failed resource but takes no recovery action.
Notify Method	Enter the full pathname of your own method to perform notification when this resource fails. This method will be called by the **server_down** event.

5. Press Enter to apply the customized resource recovery action.
6. If you use the **Notify Method**, make sure it is on all nodes in the resource group nodelist.
7. Verify and synchronize the cluster.

Note on the Fallover Option and Resource Group Availability

Be aware that if you select the **fallover** option of customized resource recovery—which could cause a resource group to migrate from its original node—the possibility exists that while the highest priority node is up, the resource group remains down. This situation occurs when an **rg_move** event moves a resource group from its highest priority node to a lower priority node, and then you stop the cluster services on the lower priority node with an option to bring the resource groups offline. Unless you bring the resource group up manually, it remains in an inactive state.

For more information on resource group availability, see the section Selective Fallover for Handling Resource Groups in Appendix B: Resource Group Behavior during Cluster Events.

Testing Customized Resource Recovery

Once you have configured the options and synchronized your cluster successfully, you are ready to test that the new options provide the desired behavior.

Testing the Fallover Action on Resource Failure

This is the default behavior. When a resource failure occurs (local_network_down or volume group quorum loss), an **rg_move** event will be run for the affected resource group. You can test this behavior by inducing a local_network_down (fail all interfaces on that network on a single node) or by inducing the LVM_SA_QUORCLOSE error (power off a disk while writes are occurring such that quorum is lost for that volume group). For additional information, see the section Quorum Issues in Appendix A: 7x24 Maintenance.

You can also use the error emulation facility to create an LVM_SA_QUORCLOSE error log record. See the HACMP Event Emulation section in Chapter 1: Troubleshooting HACMP Clusters in the *Troubleshooting Guide*.

Testing the Notify Action on Resource Failure

Induce the same failures mentioned above, selecting **notify** but no **Notify Method**. Instead of an **rg_move** event, a **server_down** event should run. Check the output in **hacmp.out**.

Testing the Notify Method

Configure a resource and resource group and specify the **notify** option for that resource, with a **Notify Method**. Induce one of the failures above to trigger the **server_down** event. The **server_down** event will call the **Notify Method** and any output from that method will be logged in **hacmp.out**.

Where You Go From Here

The next step is to configure the resource groups for the cluster, configure dependencies between resource groups, if needed, and add the resources to the resource groups. See Chapter 5: Configuring HACMP Resource Groups (Extended).

Chapter 5: Configuring HACMP Resource Groups (Extended)

This chapter describes all the options for configuring resource groups using the SMIT **Extended Configuration** path. It also contains sections for NFS considerations and for using the forced **varyon** option.

The main sections in this chapter include:

- Overview
- Configuring Resource Groups
- Configuring Resource Group Runtime Policies
- Configuring Dependencies between Resource Groups
- Adding Resources and Attributes to Resource Groups Using the Extended Path
- Customizing Inter-Site Resource Group Recovery
- Reliable NFS Function
- Forcing a Varyon of Volume Groups.

Note: Starting with HACMP 5.3, you can configure several types of dependencies between resource groups. See Configuring Dependencies between Resource Groups for information.

Overview

You may have already used the **Standard Configuration** panels to configure some resources and groups automatically. Use the **Extended Configuration** SMIT panels to add more resources and groups, to make changes, or to add more extensive customization.

The Extended Resources Configuration path includes three main menus that contain sub menus:

- HACMP Extended Resources Configuration. See Chapter 4: Configuring HACMP Cluster Topology and Resources (Extended) for this information. The SMIT panel for customizing cross-site recovery of resource groups is located here.
  - Resource Group Runtime Policies Configuration
  - Dependencies between Resource Groups
  - Workload Manager
  - Configure Resource Group Processing Order
  - Delayed Fallback Timer
  - Settling Time
  - Node Distribution Policy

- HACMP Extended Resource Group Configuration:
  - Add a Resource group
  - Change/Show a Resource group
  - Change /Show Resources and Attributes for a Resource Group
  - Remove a Resource Group
  - Show All Resources by Node or Resource Group

The chapter also includes sections on two specific configuration issues:

- Reliable NFS Function
- Forcing a Varyon of Volume Groups

Note: You can use either ASCII SMIT or WebSMIT to configure and manage the cluster. For more information on WebSMIT, see Chapter 2: Administering a Cluster Using WebSMIT.

Configuring Resource Groups

You can add resource groups with different startup, fallover and fallback policies. See the *Concepts Guide* for definitions and an overview of the different resource group policies. For information about planning new resource groups, see the *Planning Guide*. For information about migration of the pre-5.2 resource groups, see the *Installation Guide.*

This chapter explains how to configure resource groups with different combinations of inter-site resource group management with startup, fallover and fallback policies, and runtime policies.

We recommend that prior to configuring resource groups, you read the planning information. For more information, see Chapter 6: Planning Resource Groups in the *Planning Guide.*

Note: You can use either ASCII SMIT or WebSMIT to configure and manage the cluster and view interactive cluster status. Starting with HACMP 5.4, you can also use WebSMIT to navigate, configure and view the status of the and graphical displays of the running cluster. For more information about WebSMIT, see Chapter 2: Administering a Cluster Using WebSMIT.

The **Extended Configuration** path enables you to specify parameters that precisely describe the resource group's behavior at startup, fallover, and fallback, including delayed fallback timers (you cannot configure the fallback timers using the **Initialization and Standard Configuration** path).

This section describes the parameters that you can specify for resource groups, provides their definitions and limitations for their use, and includes rules on how to set these parameters correctly:

- Limitations and Prerequisites for Configuring Resource Groups
- Steps for Configuring Resource Groups in SMIT
- Dynamic Node Priority Policies

- Configuring Resource Group Runtime Policies
- Defining Delayed Fallback Timers
- Configuring Delayed Fallback Timers in SMIT
- Assigning a Delayed Fallback Policy to a Resource Group
- Adding Resources and Attributes to Resource Groups Using the Extended Path.

Limitations and Prerequisites for Configuring Resource Groups

When configuring a resource group, the following conditions apply:

- Networks can be configured to use IPAT via IP Aliases or IPAT via IP Replacement. You can add both aliased and non-aliased service IP labels as resources to resource groups.
- By default, HACMP processes resource groups in parallel. You may include a resource group in a list of resource groups that are processed serially. However, if you do not include a resource group in a serially-processed list, but specify a settling time or a delayed fallback timer for a resource group, the acquisition of this resource group is delayed. For complete information, see the section Configuring Processing Order for Resource Groups.
- Clocks on all nodes must be synchronized for the settings for fallover and fallback of a resource group to work as expected.
- You can configure sites for all resource groups. In SMIT, select the inter-site management policy according to your requirements. For planning inter-site management policies, see the *Planning Guide*.
- To view the information about resource groups and for troubleshooting purposes, use the **clRGinfo** command. Also, for troubleshooting purposes, you can use the **Show All Resources by Node or Resource Group** SMIT option.

Steps for Configuring Resource Groups in SMIT

Steps to configure a resource group:

Step	What you do...
1	Configure the **Runtime Policies** that you want to assign to your resource group. a) *(Optional.)* Configure a delayed fallback timer. See the section Defining Delayed Fallback Timers for instructions. After you have configured a delayed fallback policy, you can assign it to the resource group by specifying the appropriate fallback policy, and by adding this policy as an attribute to your resource group. b) *(Optional.)* Configure a settling time. See Configuring Resource Group Runtime Policies for instructions.
2	Define a startup policy for the resource group. Select the SMIT option **Startup Policy**. Instructions follow this list of steps.

Step	What you do...
3	Define a fallover policy for the resource group. Select the SMIT option **Fallover Policy**. Instructions follow this list of steps.
4	Define a fallback policy for the resource group. Select the SMIT option **Fallback Policy**. Instructions follow this list of steps.
5	After you have set the resource group's startup, fallover, and fallback policies as necessary, add resources and attributes to the resource group. If you have configured a delayed fallback timer, you can include it as an attribute of the resource group. If you want to use one of the predefined dynamic node priority policies, include it. Also, if you have specified the settling time, it will be used for this resource group if the appropriate startup policy is specified. If you configured the node distribution policy for the cluster, it will be used for those resource groups that have the Online Using Node Distribution Policy startup policy specified. Note, during this step you can also define a dependency between resource groups, and a customized serial processing order of resource groups, if needed.

To configure a resource group:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Resource Configuration > HACMP Extended Resource Group Configuration > Add a Resource Group** and press Enter.
3. Enter the field values as follows:

Resource Group Name	The name of the resource group must be unique within the cluster and distinct from the volume group and service IP label; it should relate to the application it serves, as well as to any corresponding device, such as `websphere_service_address`. Use no more than a total of 32 alphanumeric characters and underscores. Do not use a leading numeric. Duplicate entries and reserved words are not allowed. See List of Reserved Words.

Inter-site Management Policy

The default is **Ignore**. Select one of these options:

- **Ignore**. If you select this option, the resource group will not have ONLINE SECONDARY instances. Use this option if you use cross-site LVM mirroring. You can also use it with HACMP/XD for Metro Mirror.
- **Prefer Primary Site**. The primary instance of the resource group is brought ONLINE on the primary site at startup, the secondary instance is started on the other site. The primary instance falls back when the primary site rejoins.
- **Online on Either Site.** During startup the primary instance of the resource group is brought ONLINE on the first node that meets the node policy criteria (either site). The secondary instance is started on the other site. The primary instance does not fall back when the original site rejoins.
- **Online on Both Sites.** During startup the resource group (node policy must be defined as Online on All Available Nodes) is brought ONLINE on both sites. There is no fallover or fallback

  Note that the resource group moves to another site *only* if no node or condition exists under which it can be brought or kept ONLINE on the site where it is currently located. The site that owns the active resource group is called the primary site.

Participating Nodes (Default Node Priority)

Enter the names of the nodes that can own or take over this resource group. Enter the node with the highest priority first, followed by the other nodes in priority order. Leave a space between node names.

NOTE: If you defined sites for the cluster, this panel is divided into **Participating Nodes from Primary Site** and **Participating Nodes from Secondary Site**.

Startup Policy

Select a value from the list that defines the startup policy of the resource group:

Online On Home Node Only. The resource group should be brought online *only* on its home (highest priority) node during the resource group startup. This requires the highest priority node to be available.

Online On First Available Node. The resource group activates on the first participating node that becomes available.

If you have configured the settling time for resource groups, it will only be used for this resource group if you use this startup policy option. For information on the settling time, see the section Configuring Resource Group Runtime Policies.

Online Using Node Distribution Policy. The resource group is brought online according to the node-based distribution policy. This policy allows only one resource group to be brought online on a node during startup.

Note: Rotating resource groups upgraded from HACMP 5.1 will have the node-based distribution policy.

For more information, see Using the Node Distribution Startup Policy.

Online On All Available Nodes. The resource group is brought online on *all* nodes.

If you select this option for the resource group, ensure that resources in this group can be brought online on multiple nodes simultaneously.

Fallover Policy	Select a value from the list that defines the fallover policy of the resource group: **Fallover To Next Priority Node In The List.** In the case of fallover, the resource group that is online on only one node at a time follows the default node priority order specified in the resource group's nodelist. **Fallover Using Dynamic Node Priority.** Select one of the predefined dynamic node priority policies. See Dynamic Node Priority Policies for more information. **Bring Offline (On Error Node Only)**. Select this option to bring a resource group offline on a node during an error condition. This option is most suitable when you want to ensure that if a particular node fails, the resource group goes offline only on that node but remains online on other nodes. Selecting this option as the fallover preference when the startup preference is not **Online On All Available Nodes** may allow resources to become unavailable during error conditions. If you do so, HACMP issues an error.
Fallback Policy	Select a value from the list that defines the fallback policy of the resource group: **Fallback To Higher Priority Node In The List.** A resource group falls back when a higher priority node joins the cluster. If you select this option, you can use the delayed fallback timer that you previously specified in the **Configure Resource Group Runtime Policies** SMIT menu. If you do not configure a delayed fallback timer, the resource group falls back immediately when a higher priority node joins the cluster. See the section Defining Delayed Fallback Timers for instructions. **Never Fallback**. A resource group does *not* fall back when a higher priority node joins the cluster.

4. Press Enter to add the resource group information to the HACMP Configuration Database.

   If, during the configuration of resource groups, you select an option that prevents high availability of a resource group, HACMP issues a warning message. Also, HACMP prevents invalid or incompatible resource group configurations.

5. Return to the **Extended Resource** configuration panel or exit SMIT.

Note: For additional information on resource group behavior in clusters with sites, see the *Planning Guide*, Chapter 6, Planning Resource Groups.

Dynamic Node Priority Policies

The default node priority policy is the order in the participating nodelist. However, you may want to have a takeover node selected dynamically, according to the value of a specific system property at the time of failure.

Dynamic node priority policies based on three RMC resource attributes are preconfigured. You can see these listed when you choose **Dynamic Node Priority** as the Fallover Policy for the resource group on the **Add a Resource Group** SMIT panel:

- **`cl_highest_free_mem`** - select the node with the highest percentage of free memory
- **`cl_highest_idle_cpu`** - select the node with the most available processor time
- **`cl_lowest_disk_busy`** - select the disk that is least busy

Note: If you have defined a resource group over multiple sites (using the HACMP/XD software) and a dynamic node priority policy is configured for the group, you will receive this warning when verification runs:

```
"Warning: Dynamic Node Priority is configured in a resource group
with nodes in more than one site. The priority calculation may
fail due to slow communication, in which case the default node
priority will be used instead."
```

Configuring Resource Group Runtime Policies

Resource Group runtime policies include:

- Dependencies between resource groups. See the section Configuring Dependencies between Resource Groups.
- Resource group processing order See the section Configuring Processing Order for Resource Groups
- Workload Manager. See the section Configuring Workload Manager
- Settling Time for resource groups. See the section Configuring Resource Groups.
- Delayed Fallback Timer for resource groups. See the section Configuring Resource Groups.
- Node distribution policy. See the section Using the Node Distribution Startup Policy.

Configuring Dependencies between Resource Groups

You can set up more complex clusters by specifying dependencies between resource groups.

Business configurations that use multi-tiered applications can utilize parent/child dependent resource groups. For example, the back end database must be online before the application server. In this case, if the database goes down and is moved to a different node, the resource

group containing the application server must be brought down and back up on any node in the cluster. For more information about examples of multi-tiered applications, see the *Concepts Guide*.

Business configurations that require different applications to run on the same node, or on different nodes can use location dependency runtime policies. See Examples of Location Dependency and Resource Group Behavior in Appendix B: Resource Group Behavior during Cluster Events for more information.

In releases prior to HACMP 5.2, support for resource group ordering and customized serial processing of resources accommodated cluster configurations where a dependency existed between applications residing in different resource groups. With customized serial processing, you can specify that a resource group is processed before another resource group, on a local node. However, it is not guaranteed that a resource group will be processed in the order specified, as this depends on other resource groups policies and conditions.

The dependencies that you configure are:

- Explicitly specified using the SMIT interface
- Established cluster-wide, not just on the local node
- Guaranteed to occur in the cluster, that is, they are not affected by the current cluster conditions.

You can configure four types of dependencies between resource groups:

- Parent/child dependency
- Online On Same Node Location Dependency
- Online On Different Nodes Location Dependency
- Online On Same Site Location Dependency.

See the *Planning Guide* for more details and examples of these types of dependencies.

Considerations for Dependencies between Resource Groups

This section lists additional considerations you may need to keep in mind when configuring resource group dependencies. These include interaction with sites, use of pre-and post-event scripts, and information about the **clRGinfo** command.

- If, prior to HACMP 5.2, you were using pre-and post-event scripts or other methods, such as resource group processing ordering to establish dependencies between the applications that are supported by your cluster environment, then these methods may no longer be needed or could be significantly simplified. For more information, see the section *Dependent Resource Groups and the Use of Pre- and Post-Event Scripts* in the *Planning Guide.*
- To obtain more granular control over the resource group movements, use the **clRGinfo -a** command to view what resource groups are going to be moved during the current cluster event. Also, use the output in the **hacmp.out** file. For more information, see the section Using Resource Groups Information Commands in Monitoring an HACMP Cluster.
- Dependencies between resource groups offer a predictable and reliable way of building clusters with multi-tiered applications. However, **node_up** processing in clusters with dependencies could take more time than in the clusters where the processing of resource

groups upon **node_up** is done in parallel. A resource group that is dependent on other resource groups cannot be started until others have been started first. The **config_too_long** warning timer for **node_up** should be adjusted large enough to allow for this.

- During verification, HACMP verifies that your configuration is valid and that application monitoring is configured.
- You can configure resource group dependencies in HACMP/XD clusters that use replicated resources for disaster recovery. However, you cannot have the combination of any non-concurrent startup policy and concurrent (Online on Both Sites) inter-site management policy. You *can* have a concurrent startup policy combined with a non-concurrent inter-site management policy.

The high-level steps required to specify resource group dependencies are described in the following sections.

Steps to Configure Dependencies between Resource Groups

This section provides a high-level outline of the steps required to configure a dependency between resource groups:

1. For each application that is going to be included in dependent resource groups, configure application servers and application monitors. For more information, see Application Monitoring for Dependent Resource Groups.
2. Create resource groups and include application servers as resources. For instructions, see Configuring Resource Groups and Adding Resources and Attributes to Resource Groups Using the Extended Path.
3. Specify a dependency between resource groups. For instructions, see Configuring Resource Groups with Dependencies.
4. Use the SMIT **Verify and Synchronize HACMP Configuration** option to guarantee the desired configuration is feasible given the dependencies specified, and ensure that all nodes in the cluster have the same view of the configuration.

Application Monitoring for Dependent Resource Groups

To ensure that the applications in the dependent resource groups start successfully, we recommend that you configure multiple application monitors.

In general, we recommend that you configure a monitor that will check the running process for an application in the child resource group, and a monitor that will check the running process for an application in the parent resource group.

For a parent resource group, it is also advisable to configure a monitor in a *startup monitoring mode* to watch the application startup.This ensures that after the parent resource group is acquired, the child resource group(s) can be also acquired successfully.

For information on monitor modes that you can specify (long-running mode, startup monitoring mode, and both), see Monitor Modes in Chapter 4: Configuring HACMP Cluster Topology and Resources (Extended).

For instructions on configuring application monitoring, see Configuring Multiple Application Monitors in Chapter 4: Configuring HACMP Cluster Topology and Resources (Extended).

Configuring Resource Groups with Dependencies

You can configure four types of dependencies between resource groups:

- Parent/child dependency
- Online On Same Node Location Dependency
- Online On Different Nodes Location Dependency
- Online On Same Site Location Dependency.

See the *Planning Guide* for more details and examples of how to use these types of dependencies.

Limitations for Combinations of Location Dependencies

The following limitations apply to configurations that combine dependencies:

- Only *one* resource group can belong to a Same Node Dependency and a Different Node Dependency at the same time
- If a resource group belongs to *both* a Same Node Dependency and a Different Node Dependency, all nodes in the Same Node Dependency set have the same Priority as the shared resource group.
- Only resource groups with the same Priority within a Different Node Dependency can participate in a Same Site Dependency.

Configuring a Parent/Child Dependency Between Resource Groups

In this type of dependency, the parent resource group must be online on any node in the cluster before a child (dependent) resource group can be activated on a node. These are the guidelines and limitations:

- A resource group can serve as both a parent and a child resource group, depending on which end of a given dependency link it is placed.
- You can specify three levels of dependencies for resource groups.
- You cannot specify circular dependencies between resource groups.
- If a child resource group cannot be acquired on a node until after its parent resource group is fully functional, the child resource group goes into an ERROR state. If you notice that a resource group is in this state, you may need to troubleshoot which resources might need to be brought online manually to resolve the resource group dependency.
- When a resource group in a parent role falls over from one node to another, the resource groups that depend on it are stopped before the parent resource group falls over, and restarted again once the parent resource group is stable again. If a parent resource group is concurrent, the child resource group(s) that depend on it are stopped and restarted again. This allows the child resource group to update its knowledge about the nodes on which the parent resource group is currently online.
- For information on dynamic reconfiguration (DARE), see Reconfiguring Resources in Clusters with Dependent Resource Groups in Chapter 14: Managing the Cluster Resources.

To configure a parent/child dependency between resource groups:

1. Enter `smit hacmp`

2. In SMIT, select **Extended Configuration > HACMP Extended Resource Configuration > Configure Resource Group Run-Time Policies > Configure Dependencies between Resource Groups** > **Configure Parent/Child Dependency >Add Parent/Child Dependency between Resource Groups** and press Enter.

   The following screen appears.

3. Fill in the fields as follows:

Parent Resource Group	Select the parent resource group from the list. The parent resource group provides services upon which another resource group depends. During resource group acquisition, HACMP acquires the parent resource group on a node *before* the child resource group is acquired.
Child Resource Group	Select the child resource group from the list and press Enter. HACMP prevents you from specifying circular dependencies. The child resource group depends on services another resource group provides. During resource group acquisition, HACMP acquires the parent resource group on a node *before* the child resource group is acquired. During release, HACMP releases the child resource group *before* the parent resource group is released.

4. Press Enter and verify the cluster.

Configuring Online on the Same Node Dependency for Resource Groups

When you configure two or more resource groups to establish a location dependency between them, they belong to a *set* for that particular dependency. The following rules and restrictions apply to the Online On Same Node Dependency set of resource groups:

- All resource groups configured as part of a given Same Node Dependency set must have the same nodelist (the same nodes in the same order).
- All non-concurrent resource groups in the Same Node Dependency set must have the same Startup/Fallover/Fallback Policies.
  - Online Using Node Distribution Policy is not allowed for Startup.
  - If a Dynamic Node Priority Policy is chosen as Fallover Policy, then all resource groups in the set must have the same policy.
  - If one resource group in the set has a fallback timer, it applies to the set.
  - All resource groups in the set must have the same setting for fallback timers.
- Both concurrent and non-concurrent resource groups are allowed.
- You can have more than one Same Node Dependency set in the cluster.
- All resource groups in the Same Node Dependency set that are active (ONLINE) are required to be ONLINE on the same node, even though some resource groups in the set may be OFFLINE or in the ERROR state.
- If one or more resource groups in the Same Node Dependency set fails, HACMP tries to place all resource groups in the set on the node that can host all resource groups that are currently ONLINE (the ones that are still active) plus one or more failed resource groups.

To configure an Online on Same Node dependency between resource groups:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > HACMP Extended Resource Configuration > Configure Resource Group Run-Time Policies > Configure Dependencies between Resource Groups** > **Configure Online on Same Node Dependency >Add Online on Same Node Dependency between Resource Groups** and press Enter.

   The following screen appears.
3. Fill in the field as follows:

Resource Groups to be Online on the same node	Select the resource groups from the list to be in this set of resource groups to be acquired and brought ONLINE on the same node (according to the startup policy and the availability of the node required). On fallback and fallover, the resource groups are processed simultaneously and brought ONLINE on the same target node (using the fallover and fallback policy defined for these groups).

4. Press Enter.
5. Verify the configuration.

Configuring Online on Different Nodes Dependency for Resource Groups

When you configure two or more resource groups to establish a location dependency between them, they belong to a *set* for that particular dependency.The following rules and restrictions apply to the Online On Different Nodes Dependency set of resource groups:

- Only one Online On Different Nodes Dependency set is allowed per cluster.
- Each resource group in the set should have a different home node for startup.
- When you configure resource groups in the Online On Different Nodes Dependency set you assign priorities to each resource group in case there is contention for a given node at any point in time. You can assign High, Intermediate, and Low priority. Higher priority resource groups take precedence over lower priority groups at startup, fallover, and fallback:
  - If a resource group with High Priority is ONLINE on a node, then no other resource group in the Different Nodes Dependency set can come ONLINE on that node.
  - If a resource group in this set is ONLINE on a node, but a resource group with a higher priority falls over or falls back to this node, the resource group with the higher priority will come ONLINE and the one with the lower priority will be taken OFFLINE and moved to another node if this is possible.
  - Resource groups with the same priority cannot come ONLINE (startup) on the same node. Priority of a resource group for a node *within the same Priority Level* is determined by alphabetical order of the groups.
  - Resource groups with the same priority do not cause one another to be moved from the node after a fallover or fallback.
  - If a parent/child dependency is specified, then the child cannot have a higher priority than its parent.

To configure an Online On Different Nodes dependency between resource groups:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > HACMP Extended Resource Configuration > Configure Resource Group Run-Time Policies > Configure Dependencies between Resource Groups** > **Configure Online on Same Node Dependency >Add Online on Different Nodes Dependency between Resource Groups** and press Enter. The following screen appears.
3. Fill in the fields as follows and then press Enter

High Priority Resource Group(s)	Select the resource groups to be in this set of resource groups to be acquired and brought ONLINE (according to the startup policy and the availability of the node required) before lower priority resource groups. On fallback and fallover, these resource groups are processed simultaneously and brought ONLINE on different target nodes before any other groups are processed. If different target nodes are unavailable for fallover or fallback, these groups (same priority level) can remain on the same node. The highest relative priority within this list is the group listed first (on the left), as for the nodelist.
Intermediate Priority Resource Group(s)	Select the resource groups to be in this set of resource groups to be acquired and brought ONLINE (according to the startup policy and the availability of the node required) after the high priority groups and before low priority resource groups are brought ONLINE. On fallback and fallover, these resource groups are processed simultaneously and brought ONLINE on different target nodes after the high priority groups and before low priority resource groups are processed. If different target nodes are unavailable for fallover or fallback, these groups (same priority level) can remain on the same node. The highest relative priority within this list is the group listed first (on the left), as for the nodelist.
Low Priority Resource Group(s)	Select the resource groups to be in this set of resource groups to be acquired and brought ONLINE (according to the startup policy and the availability of the node required) after the higher priority resource groups are brought ONLINE. On fallback and fallover, these resource groups are brought ONLINE on different target nodes after the higher priority groups are processed. Higher priority groups moving to a node may cause these groups to be moved or taken OFFLINE.

4. Continue configuring runtime policies for other resource groups or verify the cluster.

Configuring Online on the Same Site Dependency for Resource Groups

When you configure two or more resource groups to establish a location dependency between them, they belong to a *set* for that particular dependency. The following rules and restrictions are applicable to Online On Same Site Dependency set of resource groups:

- All resource groups in a Same Site Dependency set must have the same Inter-site Management Policy but may have different Startup/Fallover/Fallback Policies. If fallback timers are used, these must be identical for all resource groups in the set.
- All resource groups in the Same Site Dependency set must be configured so that the nodes that can own the resource groups are assigned to the same primary and secondary sites.
- Online Using Node Distribution Policy Startup Policy is supported.
- Both concurrent and non-concurrent resource groups are allowed.
- You can have more than one Same Site Dependency set in the cluster.
- All resource groups in the Same Site Dependency set that are active (ONLINE) are required to be ONLINE on the same site, even though some resource groups in the set may be OFFLINE or in the ERROR state.
- If you add a resource group included in a Same Node Dependency set to a Same Site Dependency set, then you must add all the other resource groups in the Same Node Dependency set to the Same Site Dependency set.

To configure an Online On Same Site Dependency between resource groups:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > HACMP Extended Resource Configuration > Configure Resource Group Run-Time Policies > Configure Dependencies between Resource Groups** > **Configure Online on Same Site Dependency >Add Online on Same Site Dependency between Resource Groups** and press Enter.

   The following screen appears.
3. Fill in the field as follows and press Enter:

Resource Groups to be Online on the same site	Select the resource groups from the list to be in this set of resource groups to be acquired and brought ONLINE on the same site (according to the site and node startup policy). On fallback and fallover, the resource groups are processed simultaneously and brought ONLINE on the same site.

4. Verify the cluster.

Configuring Processing Order for Resource Groups

This section describes how to set up the order in which HACMP acquires and releases resource groups.

By default, HACMP acquires and releases resource groups in parallel. If you upgraded your cluster from previous releases, see the *Planning Guide* for more information on which processing order is used for various resource groups in this case.

Resource groups acquisition occurs in the following order:

1. Those resource groups for which the customized order is specified are acquired in the customized serial order.
2. If some of the resource groups in the cluster have dependencies between them, these resource groups are acquired in phases. Parent resource groups are acquired before the child resource groups and resource group location dependencies are taken into account.
3. Resource groups which must mount NFS only are processed in the specified order.
4. Resource groups which are not included in the customized ordering lists are acquired in parallel.

Resource groups release occurs in the following order:

1. Those resource groups for which no customized order have been specified are released in parallel.
2. HACMP releases resource groups that are included in the customized release ordering list.
3. If some of the resource groups in the cluster have dependencies between them, these resource groups are released in phases. Child resource groups are released before the parent resource groups are released, for example.
4. Resource groups which must unmount NFS are processed in the specified order.

Resource Groups Processing Order and Timers

HACMP acquires resource groups in parallel, but if the settling time or the delayed fallback timer policy is configured for a particular resource group, HACMP delays its acquisition for the duration specified in the timer policy.

Settling and delayed fallback timers do not affect the release process.

Prerequisites and Notes

The following sections detail limitations of the resource group ordering.

Serial Processing Notes

When you configure individual resource groups that depend on other resource groups, you can customize to use the serial processing order that will dictate the order of processing on the local node. If you specify dependencies between resource groups, the order in which HACMP processes resource groups cluster-wide is dictated by the dependency.

- Specify the same customized serial processing order on all nodes in the cluster. To do this, you specify the order on one node and synchronize cluster resources to propagate the change to the other nodes in the cluster. Also, since resource group dependencies also override any serial processing order, make sure that the serial order you specify does not contradict the dependencies. If it does, it will be ignored.
- If you have specified serial processing order for resource groups, and if in some of the resource groups only the NFS cross-mounting takes place during the acquisition (**node_up** event), or release (**node_down** event), then HACMP automatically processes these resource groups *after* other resource groups in the list.

- If you remove a resource group that has been included in the customized serial ordering list from the cluster, then the name of that resource group is automatically removed from the processing order list. If you change a name of a resource group, the list is updated appropriately.

Parallel Processing Notes

In clusters where some groups have dependencies defined, these resource groups are processed in parallel using event phasing. For information on the order of processing in clusters with dependent resource groups, see the Job Types: Processing in Clusters with Dependent Resource Groups section in Chapter 2: Using Cluster Log Files in the *Troubleshooting Guide*.

If, prior to migrating to HACMP 5.3 and up (where parallel processing is the default) you had pre- and post-event scripts configured for specific cluster events, you may need to change them as they may no longer work as expected. See if you can reconfigure the resource groups to take advantage of the new dependency options. If you want to continue using these scripts for these resource groups, make sure you add each of the resource groups to the serial processing lists in SMIT. For more information, see the section Resource Groups Processed in Parallel and the Use of Pre- and Post-Event Scripts in Chapter 7: Planning for Cluster Events, in the *Planning Guide.*

Error Handling

If an error occurs during the acquisition of a resource group, recovery procedures are run *after* the processing of all other resource groups is complete. See the section Updated Cluster Event Processing in Chapter 7: Planning for Cluster Events, in the *Planning Guide*.

If an error occurs during the release of a resource group, the resource group goes offline temporarily while HACMP tries to recover it. If it moves to the ERROR state, you should take care of it manually.

Steps for Changing Resource Group Processing Order

To view or change the current resource group processing order in SMIT:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Resource Configuration > Configure Resource Group Run-time Policies > Configure Resource Group Processing Ordering** and press Enter.

   SMIT displays the current processing order for the resource groups.
3. Enter field values as follows:

Resource Groups Acquired in Parallel	The current list of resource groups which are acquired in parallel by HACMP on this node.
Serial Acquisition Order	The current serial order in which HACMP serially acquires the specified resource groups on this node.

New Serial Acquisition Order	Enter the new list of resource group names. This list is the new serial order in which you want HACMP to acquire the specified resource groups on this cluster node. The resource groups that are not included in this list are acquired in parallel by default.
Resource Groups Released in Parallel	The current list of resource groups which are released in parallel by HACMP on this node.
Serial Release Order	The current serial order in which HACMP releases these resource groups on this node.
New Serial Release Order	Enter the new list of resource group names. This list is the new serial order in which you want HACMP to release the specified resource groups on this cluster node.The resource groups that are not included in this list are released in parallel by default.

4. Press Enter to accept the changes. HACMP checks that a resource group name is entered only once on a list and that all specified resource groups are configured in the cluster. Then it stores the changes in the HACMP Configuration Database.
5. Synchronize the cluster in order for the changes to take effect across the cluster.

You can determine whether or not the resource groups are being processed in the expected order based on the content of the event summaries. For more information, see the section Tracking Parallel and Serial Processing of Resource Groups in the **hacmp.out** File in Chapter 2: Using Cluster Log Files in the *Troubleshooting Guide.*

Configuring Workload Manager

IBM offers AIX 5L Workload Manager (WLM) as a system administration resource included with AIX 5L. WLM allows users to set targets for and limits on CPU time, physical memory usage, and disk I/O bandwidth for different processes and applications; this provides better control over the usage of critical system resources at peak loads. HACMP allows you to configure WLM classes in HACMP resource groups so that the starting, stopping, and active configuration of WLM can be under cluster control.

For complete information on how to set up and use Workload Manager, see the *AIX 5L Workload Manager (WLM)* Redbook at the URL:

http://www.ibm.com/redbooks

Steps for Configuring WLM in HACMP

Configuring WLM classes in HACMP involves these basic steps:

1. Configure WLM classes and rules, using the appropriate AIX 5L SMIT panels, as described below.
2. If you select a configuration other than the default (“HACMP_WLM_config”), specify the WLM configuration to be used in HACMP, as described below.

3. Assign the classes for this configuration to a resource group, selecting from a picklist of the classes associated with the default WLM configuration or the configuration you specified in Step 2. For instructions on adding resources to resource groups, see Adding Resources and Attributes to Resource Groups Using the Extended Path in this chapter.
4. After adding the WLM classes to the resource group—or after all resource group configuration is complete—verify and synchronize the configuration.

Note: Once WLM is configured in HACMP, HACMP starts and stops WLM. If WLM is already running when HACMP is started, HACMP restarts it with a new configuration file. Therefore, only the WLM rules associated with classes in a resource group that can be acquired on a given node will be active on that node. Once HACMP is stopped, WLM will be switched back to the configuration it was using when it was started.

Creating a New Workload Manager Configuration

To set up WLM classes and rules, use the AIX 5L SMIT panels.

1. In AIX 5L SMIT, select **Performance & Resource Scheduling** > **Workload Management** > **Work on alternate configurations > Create a configuration.** (You can also get to the "alternate configurations" panel by typing `smitty wlm`.)
2. Enter the new name for the configuration in the **New configuration name** field. It is recommended to use the default name that HACMP supplies: HACMP_WLM_config.
3. Define classes and rules for the HACMP configuration.

Defining a Non-Default Workload Manager Configuration in HACMP

You may have a non-default Workload Manager configuration. In this case, make this configuration known to HACMP, so that it is managed.

To ensure that a non-default Workload Manager Configuration is managed by HACMP:

1. Change the WLM runtime parameters to specify the HACMP configuration.
2. From the main HACMP SMIT panel, select **Extended Configuration > Extended Resource Configuration > Configure Resource Group Run-time Policies > Configure HACMP Workload Manager Parameters**.

   This field indicates the WLM configuration to be managed by HACMP. By default, the configuration name is set to **HACMP_WLM_config**.
3. Specify a different configuration name if needed.

Verification of the Workload Manager Configuration

After adding WLM classes to resource groups, or after you have finished configuring all your resource groups, verify that the configuration is correct. The verification step is included in the synchronization process, as described in Synchronizing Cluster Resources later in this chapter.

Verification checks for the following conditions:

- For each resource group with which a WLM class is associated, an application server is associated with this resource group. It is not required that an application server exists in the resource group, but it is expected. HACMP issues a warning if no application server is found.
- Each WLM class defined to an HACMP resource group exists in the specified HACMP WLM configuration directory.
- A non-concurrent resource group (that does not have the Online Using Node Distribution Policy startup policy) does not contain a secondary WLM class without a primary class.
- A resource group with the startup policy Online on All Available Nodes has only a primary WLM class.
- A resource group with the startup policy Online Using Node Distribution Policy has only a primary WLM class.

Note: The verification utility cannot check class assignment rules to verify that the correct assignment will take place, since HACMP has no way of determining the eventual gid, uid and pathname of the user application. The user is entirely responsible for assigning user applications to the WLM classes when configuring WLM class assignment rules.

Cluster verification looks only for obvious problems and cannot verify all aspects of your WLM configuration; for proper integration of WLM with HACMP, you should take the time to plan your WLM configuration carefully in advance.

Reconfiguration, Startup, and Shutdown of WLM by HACMP

This section describes the way WLM is reconfigured or started or stopped once you have placed WLM under the control of HACMP.

Workload Manager Reconfiguration

When WLM classes are added to an HACMP resource group, then at the time of cluster synchronization on the node, HACMP reconfigures WLM so that it will use the rules required by the classes associated with the node. In the event of dynamic resource reconfiguration on the node, WLM will be reconfigured in accordance with any changes made to WLM classes associated with a resource group.

Workload Manager Startup

WLM startup occurs either when the node joins the cluster or when a dynamic reconfiguration of the WLM configuration takes place.

The configuration is node-specific and depends upon the resource groups in which the node participates. If the node cannot acquire any resource groups associated with WLM classes, WLM will not be started.

For a non-concurrent resource group with the startup policy other than Online Using Node Distribution Policy, the startup script will determine whether the resource group is running on a primary or on a secondary node and will add the corresponding WLM class assignment rules to the WLM configuration.

For each concurrent access resource group, and for each non-concurrent resource group with the startup policy Online Using Node Distribution Policy that the node can acquire, the primary WLM class associated with the resource group will be placed in the WLM configuration; the corresponding rules will be put into the rules table.

Finally, if WLM is currently running and was not started by HACMP, the startup script restarts WLM from the user-specified configuration, saving the prior configuration. When HACMP is stopped, it returns WLM back to its prior configuration.

Failure to start up WLM generates an error message logged in the **hacmp.out** log file, but node startup and/or the resource reconfiguration will proceed normally.

Workload Manager Shutdown

WLM shutdown occurs either when the node leaves the cluster or on dynamic cluster reconfiguration. If WLM is currently running, the shutdown script checks if the WLM was running prior to being started by the HACMP and what configuration it was using. It then either does nothing (if WLM is not currently running), or stops WLM (if it was not running prior to HACMP startup), or stops it and restarts it in the previous configuration (if WLM was Configuring Resources in a Resource Group

Once you have defined a resource group, you assign resources to it. SMIT cannot list possible shared resources for the node (making configuration errors more likely) if the node is powered off.

Configuring a Settling Time for Resource Groups

The settling time specifies how long HACMP waits for a higher priority node (to join the cluster) to activate a resource group that is currently offline on that node. If you set the settling time, HACMP waits for the duration of the settling time interval to see if a higher priority node may join the cluster, rather than simply activating the resource group on the first possible node that reintegrates into the cluster.

To configure a settling time for resource groups, do the following:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Resource Configuration > Configure Resource Group Run-Time Policies > Configure Settling Time for Resource Group** and press Enter.

   The **Configure Settling Time** panel appears.

3. Enter field values as follows:

Settling Time (in Seconds)	Enter any positive integer number in this field. The default is zero. In this case the resource group does not wait before attempting to start on a joining higher priority node. If you set the settling time, then if the currently available node that reintegrated into the cluster is not the highest priority node, the resource group waits for the duration of the settling time interval. When the settling time expires, the resource group is acquired on the node which has the highest priority among the list of nodes that joined the cluster during the settling time interval. If no nodes joined the cluster, the resource group remains offline. The settling time is only valid for resource groups that have the **Online on First Available Node** startup policy.

4. Press Enter to commit the changes and synchronize the cluster. This settling time is assigned to all resource groups with the **Online on First Available Node** startup policy.

You can change, show or delete a previously configured settling time using the same SMIT path as described for configuring a settling time.

For an example of the event summary showing the settling time, see the Event Summary for the Settling Time section in Chapter 2: Using Cluster Log Files in the *Troubleshooting Guide*.

Defining Delayed Fallback Timers

A delayed fallback timer lets a resource group fall back to its higher priority node at a specified time. This lets you plan for outages for maintenance associated with this resource group.

You can specify a recurring time at which a resource group will be scheduled to fall back, or a specific time and date when you want to schedule a fallback to occur.

You can specify the following types of delayed fallback timers for a resource group:

- Daily
- Weekly
- Monthly
- Yearly
- On a specific date.

Note: It is assumed that the delayed timer is configured so that the fallback time is valid. If the configured time occurs in the past or is not valid, you receive a warning and the delayed fallback policy is ignored. If you use a specific date, the fallback attempt is made only once, at the specified time.

To make a resource group use a delayed fallback policy, follow these steps:

Step	What you do...
1	Configure a delayed fallback timer that you want to use. After you have configured the delayed fallback timer, you can use it in one or several resource groups as the default fallback policy. For instructions, see the following section.
2	Select the **Fallback to Higher Priority Node** option from the picklist of fallback policies for your resource group. You can do so when configuring a resource group. For instructions, see the section Steps for Configuring Resource Groups in SMIT.
3	Assign a fallback timer to a resource group, by adding it as an attribute to the resource group. If the **delayed fallback timer** entry does not show up in the list of attributes/resources that you can add to a resource group, this indicates that you did not follow the instructions in steps 1 and 2, because HACMP only displays attributes and resources that are valid in each particular case. For instructions, see the section Assigning a Delayed Fallback Policy to a Resource Group.

Configuring Delayed Fallback Timers in SMIT

To configure a delayed fallback timer, follow these steps:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Resource Configuration > Configure Resource Group Run-Time Policies > Configure Delayed Fallback Timer Policies > Add a Delayed Fallback Timer Policy** and press Enter.

   A picklist **Recurrence for Fallback Timer** displays. It lists **Daily, Weekly, Monthly**, **Yearly** and **Specific Date** policies.
3. Select the timer policy from the picklist and press Enter. Depending on which option you select, a corresponding SMIT panel displays that lets you configure this type of a fallback policy.

Assigning a Delayed Fallback Policy to a Resource Group

You must define the delayed fallback policies before you can assign them as attributes to resource groups.

To assign a delayed fallback policy to a resource group:

1. In HACMP SMIT, create a resource group, or select an existing resource group.
2. In SMIT, select **Extended Configuration > Extended Resource Configuration > HACMP Extended Resource Group Configuration > Change/Show Resource and Attributes for a Resource Group** and press Enter. SMIT displays a list of resource groups.
3. Select the resource group for which you want to assign a delayed fallback policy. The following panel appears. (The SMIT panel is abbreviated below. All valid options for the resource group are displayed based on the startup, fallover and fallback preferences that you have specified for the resource group.)

4. Enter field values as follows:

Resource Group Name	The name of the selected resource group displays here.
Inter-site Management Policy	**Ignore** (default) is used for most resource groups.
Participating Node Names (Default Node Priority)	The names of the nodes that can own or take over this resource group. The node with the highest priority is listed first, followed by the nodes with the lower priorities.
Dynamic Node Priority (Overrides default)	The default is blank (the ordered nodelist). The preconfigured policies are listed. Note that this SMIT option displays only if you have previously selected **Fallover Using Dynamic Node Priority** as a fallover policy for this resource group.
Fallback Timer Policy (empty is immediate)	The default is blank (the resource group falls back immediately after a higher priority node joins). All configured fallback timer policies are listed in the picklist. Note that this SMIT option displays only if you have previously selected **Fallback to Higher Priority Node in the List** as a fallback policy for this resource group.

5. Press the F4 key to see the picklist in the **Fallback Timer Policy** field and select the fallback timer policy you want to use for this resource group.
6. Press Enter to commit the changes. The configuration is checked before populating the HACMP Configuration Database. You can assign the same fallback timer policy to other resource groups.
7. Assign fallback timer policies to other resource groups and synchronize the cluster when you are done.

Using the Node Distribution Startup Policy

For each resource group in the cluster, you can specify a startup policy to be Online Using Node Distribution Policy. The only distribution policy supported in HACMP 5.3 and up is node-based distribution. You can use this policy whether or not you have sites configured in the cluster.

Note: If you upgrade to HACMP 5.3 or 5.4 from a previous release that allowed network-based distribution, that configuration is automatically changed to node-based distribution.

This distribution policy is a cluster-wide attribute that causes the resource groups to distribute themselves in a way that only one resource group is acquired on a node during startup. Using this policy ensures that you distribute your CPU-intensive applications on different nodes.

If two or more resource groups are offline at the time when a particular node joins, the node acquires the resource group that has the least number of nodes in its nodelist. After considering the number of nodes, HACMP sorts the list of resource groups alphabetically.

Note: If one of the resource groups is a parent resource group (has a dependent resource group), HACMP gives preference to the parent resource group.

Prerequisites and Limitations for the Node Distribution Startup Policy

When configuring the node distribution startup policy, take into consideration the following:

- If the number of resource groups is larger than the number of cluster nodes, HACMP issues a warning. It is recommended that all resource groups that use node-based distribution have potential nodes on which they could be brought online during the cluster startup.
- Resource groups configured for distribution during startup cannot have the fallover policy set to Bring Offline (on Error Node Only). If you select this combination of policies, HACMP issues an error.
- Resource groups configured for distribution during startup must use the Never Fallback policy. This is the only fallback policy HACMP allows for such resource groups.
- If you configure multiple resource groups to use the Online Using Node Distribution startup policy, and you select the Prefer Primary Site inter-site management policy for all groups, the node-based distribution policy ensures that the primary site hosts *one group per node*. Whether the resource group will fall back to the primary site depends on the availability of nodes on that site.

HACMP allows only valid startup, fallover and fallback policy combinations and prevents you from configuring invalid combinations.

Adding Resources and Attributes to Resource Groups Using the Extended Path

Keep the following in mind as you prepare to define the resources in your resource group:

- If you are configuring a resource group, first you configure timers (optional), startup, fallover, and fallback policies for a resource group, and then add specific resources to it. For information on configuring resource groups, see Configuring Resource Groups.
- You cannot change a resource group's policies once it contains resources. If you have added resources, you need to remove them prior to changing the resource group's policies.
- If you configure a non_concurrent resource group (with the Online on Home Node startup policy) with an NFS mount point, you must also configure the resource to use IP Address Takeover. If you do not do this, takeover results are unpredictable.You should also set the field value **Filesystems Mounted Before IP Configured** to **true** so that the takeover process proceeds correctly.
- A resource group may include multiple service IP addresses. When a resource group configured with IPAT via IP Aliasing is moved, all service labels in the resource group are moved as aliases to the available interfaces, according to the resource group management

policies in HACMP. For more information on how HACMP handles the resource groups configured with IPAT via IP Aliasing see Appendix B: Resource Group Behavior during Cluster Events.

- When setting up a non-concurrent resource group with the startup policy of either Online on Home Node Only or Online on First Available Node, and with an IPAT via IP Address Replacement configuration, each cluster node should be configured in no more than (N + 1) resource groups on a particular network. Here, N is the number of backup (standby) interfaces on a particular node and network.
- IPAT functionality is not applicable to concurrent resource groups.
- If you configure application monitoring, remember that HACMP can monitor only one application in a given resource group, so you should put applications you intend to have HACMP monitor in separate resource groups.
- If you plan to request HACMP to use a forced varyon option to activate volume groups in case a normal varyon operation fails due to a loss of quorum, the logical volumes should be mirrored. It is recommended to use the **super strict** disk allocation policy for the logical volumes in AIX 5L.

Steps for Adding Resources and Attributes to Resource Groups (Extended Path)

To configure resources and attributes for a resource group:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Resource Configuration > Extended Resource Group Configuration** > **Change/Show Resources and Attributes for a Resource Group** and press Enter.

   SMIT displays a list of defined resource groups.
3. Select the resource group you want to configure and press Enter. SMIT returns the panel that matches the type of resource group you selected, with the **Resource Group Name**, **Inter-site Management Policy**, and **Participating Node Names (Default Node Priority)** fields filled in.

   SMIT displays only valid choices for resources, depending on the resource group startup, fallover, and fallback policies that you selected.

   Note: Be aware that once you add resources to a resource group, its startup, fallover, and fallback policies cannot be changed unless you remove the resources. You can only change the resource group policies in a resource group that does not contain any resources yet. Plan your resource group's policies in advance, before adding resources to it.

   If the participating nodes are powered on, press F4 to list the shared resources. If a resource group/node relationship has not been defined, or if a node is not powered on, the picklist displays the appropriate warnings.

4. Enter the field values as follows (non-concurrent resource group is shown):

Dynamic Node Priority (Overrides default)	Select the dynamic node priority policy.The default is blank (the ordered nodelist). The preconfigured dynamic node priority policies are listed.
Service IP Labels/Addresses	Enter the IP labels/addresses to be taken over when this resource group is taken over, or select an IP label/address from the picklist. The picklist includes IP labels/addresses which rotate or may be taken over.
Application Servers	Enter or select from the picklist the application servers to include in the resource group.
Volume Groups	Identify the shared volume groups that should be varied on when this resource group is acquired or taken over. Select the volume groups from the picklist or enter desired volume groups names in this field. If you previously requested that HACMP collect information about the appropriate volume groups, the picklist displays list of all shared volume groups in the resource group *and* the volume groups that are currently available for import onto the resource group nodes. Specify the shared volume groups in this field if you want to leave the field **Filesystems (empty is All for specified VGs)** blank *and* to mount all filesystems in the volume group. If you specify more than one volume group in this field, then all filesystems in all specified volume groups will be mounted; you cannot choose to mount all filesystems in one volume group and not to mount them in another. For example, in a resource group with two volume groups, vg1 and vg2, if the **Filesystems (empty is All for specified VGs)** is left blank, then all the filesystems in vg1 and vg2 will be mounted when the resource group is brought up. However, if the **Filesystems (empty is All for specified VGs)** has only filesystems that are part of the vg1 volume group, then none of the filesystems in vg2 will be mounted, because they were not entered in the **Filesystems (empty is All for specified VGs)** field along with the filesystems from vg1. If you have previously entered values in the **Filesystems** field, the appropriate volume groups are already known to the HACMP software.

Use Forced Varyon of Volume Groups, if Necessary	The default is **false**. If this flag is set to **true**, HACMP uses a forced varyon to bring each volume group that belongs to this resource group online in the event that a normal varyon for the volume group fails due to a lack of quorum, *and* if HACMP finds at least one complete copy of every logical partition in every logical volume available for this volume group. Use this option *only* for volume groups in which every logical volume is mirrored. It is recommended to use the **super strict** disk allocation policy; the forced varyon operation is unlikely to be successful for other choices of logical volume configuration.
Filesystems (empty is All for specified VGs)	Leave this field blank, if you want *all* filesystems in the specified volume groups to be mounted by default when the resource group, containing this volume group, is brought online. If you leave the **Filesystems (empty is All for specified VGs)** field blank *and* specify the shared volume groups in the **Volume Groups** field below, all filesystems will be mounted in the volume group. If you leave the **Filesystems** field blank and do not specify the volume groups in the field below, no filesystems will be mounted. You may also select individual filesystems to include in the resource group. Press F4 to see a list of the filesystems. In this case, only the specified filesystems will be mounted when the resource group is brought online. **Filesystems (empty is All for specified VGs)** is a valid option *only* for non-concurrent resource groups.
Filesystems Consistency Check	Identify the method to check consistency of filesystems, **fsck** (default) or **logredo** (for fast recovery).
Filesystems Recovery Method	Identify the recovery method for the filesystems, parallel (for fast recovery) or sequential (default). Do *not* set this field to **parallel** if you have shared, nested filesystems. These must be recovered sequentially. (Note that the cluster verification process does not report filesystem and fast recovery inconsistencies.)

Filesystems Mounted Before IP Configured	Specify whether, on takeover, HACMP takes over volume groups and mounts a failed node's filesystems before or after taking over the failed node's IP address or addresses. The default is **false**, meaning the IP address is taken over first. Similarly, upon reintegration of a node, the IP address is acquired before the filesystems. Set this field to **true** if the resource group contains filesystems to export. This is so that the filesystems will be available once NFS requests are received on the service address.
Filesystems/Directories to Export	Identify or select from the picklist the filesystems or directories to be NFS exported. The filesystems should be a subset of the filesystems listed in **Filesystems** fields above. The directories should be contained in one of the filesystems listed above. (Example: `/fs1`)
Filesystems/Directories to NFS Mount	Identify the filesystems or directories to NFS mount. All nodes in the resource chain will attempt to NFS mount these filesystems or directories while the owner node is active in the cluster.Example: Using `/fs1` from the previous entry, you enter the remote mount, then the local mount: `/rfs1;/fs1`.
Network for NFS Mount	(*Optional*.) Select the network where you want to NFS mount the filesystems from a picklist of a previously defined IP networks. This field is relevant only if you have filled in the **Filesystems/Directories to NFS Mount** field. The **Service IP Labels/IP Addresses** field should contain a service label which is on the network you select. **Note:** You can specify more than one service label in the **Service IP Labels/IP Addresses field**. It is highly recommended that at least one entry be an IP label on the network chosen here. If the network you have specified is unavailable when the node is attempting to NFS mount, it will seek other defined, available IP networks in the cluster on which to establish the NFS mount.
Raw Disk PVIDs	Press F4 for a listing of the PVIDs and associated hdisk device names. If you have previously entered values in the **Filesystems** or **Volume groups** fields, the appropriate disks are already known to the HACMP software. If you are using an application that directly accesses raw disks, list the raw disks here.

Tape Resources	Enter or select from the picklist the tape resources that you want started on the resource group. The picklist displays a list of resources previously defined in the **Define Tape Resources** panel.
Fast Connect Services	Press F4 to select from a list of Fast Connect resources common to all nodes in the resource group, specified during the initial configuration of Fast Connect. If you are adding Fast Connect fileshares, make sure you have defined their filesystems in the resource group.
Communication Links	Enter the communication links (defined previously in the **Configure Communication Adapters and Links** SMIT panels) to be started by HACMP. Press F4 to see a list of defined communication links. If adding **SNA-over-LAN** links, make sure you have also added a service IP label to the resource group.
Miscellaneous Data	Miscellaneous Data is a string placed into the MISC_DATA environment variable. The MISC_DATA environment variable is accessible by scripts, for example pre- and post-event scripts and application server start and stop scripts.
Primary Workload Manager Class	Select from the picklist of Workload Manager (WLM) classes associated with the HACMP WLM configuration specified. • For non-concurrent resource groups with the startup policy of Online on Home Node Only, or Online on First Available Node, if no secondary WLM class is specified, all nodes use the primary WLM class. If a secondary class is specified, only the primary node uses the primary WLM class. • For non-concurrent resource groups with the startup policy of Online Using a Distribution Policy, all nodes in the resource group will use the primary WLM class. • For concurrent resource groups, all nodes in the resource group will use the primary WLM class.

Secondary Workload Manager Class	(*Optional*) Press F4 and select from the picklist of Workload Manager class associated with this resource group. Only non-concurrent resource groups with the startup policy of either Online On Home Node Only, or Online on First Available node are allowed to use secondary WLM classes. If no secondary WLM class is specified, all nodes in the resource group use the primary WLM class. If you specify a secondary class here, the primary node uses the primary WLM class and all other nodes use the secondary WLM class.
Automatically Import Volume Groups	Specifies whether HACMP should *automatically* import those volume groups that are defined in the **Volume Groups** or **Concurrent Volume Groups** fields. By default, **Automatically Import Volume Groups** flag is set to **false**. If **Automatically Import Volume Groups** is set to **false**, then selected volume groups will not be imported automatically. In this case, when you add volume groups to the resource group, make sure that the selected volume groups have already been imported to each of the nodes using the **importvg** command or C-SPOC. If **Automatically Import Volume Groups** is set to **true**, then when you press Enter, HACMP determines whether the volume group that you entered or selected in the **Volume Groups** or **Concurrent Volume Groups** fields needs to be imported to any of the nodes in the resource group, and automatically imports it, if needed.
Fallback Timer Policy (empty is immediate)	This field displays only if you have previously selected **Fallback to Higher Priority Node in the List** as a fallback policy. The default is blank (the resource group falls back immediately after a higher priority node joins). The picklist contains all configured fallback timer policies.

5. Press Enter to add the values to the HACMP Configuration Database.
6. Return to the top of the **Extended Configuration** menu and synchronize the cluster.

Customizing Inter-Site Resource Group Recovery

When you install HACMP 5.3 or 5.4 and configure a new cluster with sites and an HACMP/XD product, selective fallover of resources included in replicated resource groups is enabled by default. If necessary for recovery, HACMP moves the resource group containing the resource to the other site.

If you have migrated from a previous release, the pre-5.3 release behavior is the default. In releases prior to 5.3, a particular instance of a resource group can fall over within one site, but cannot move between sites. If no nodes are available on the site where the affected instance resides, that instance goes into ERROR or ERROR_SECONDARY state. It does not stay on the node where it failed. This behavior applies to both primary and secondary instances.

Note that in HACMP 5.3 and 5.4, as in prior releases, even if the Cluster Manager cannot initiate a selective fallover inter-site **rg_move** (if this recovery is disabled), it will still move the resource group if a **node_down** or **node_up** event occurs, and you can manually move the resource group across sites.

Enabling or Disabling Selective Fallover between Sites

You can change the resource group recovery policy to allow or disallow the Cluster Manager to move a resource group to another site in cases where it can use selective fallover to avoid having the resource group go into ERROR state.

Inter-Site Recovery of Both Instances of Replicated Resource Groups

If selective fallover across sites is enabled, HACMP tries to recover both the primary and the secondary instance of a resource group:

- If an acquisition failure occurs while the secondary instance of a resource group is being acquired, the Cluster Manager tries to recover the resource group's secondary instance, as it does for the primary instance. If no nodes are available for the acquisition, the resource group's secondary instance goes into global ERROR_SECONDARY state.
- If quorum loss is triggered, and the resource group has its secondary instance online on the affected node, HACMP tries to recover the secondary instance on another available node.
- If a **local_network_down** occurs on an **XD_data** network, HACMP moves replicated resource groups that are ONLINE on the particular node that have GLVM or HAGEO resources to another available node on that site. This functionality of the primary instance is mirrored to the secondary instance so that secondary instances may be recovered via selective fallover. (For more information on XD_data networks and GLVM, see the *HACMP/XD for GLVM: Planning and Administration Guide*).

Using SMIT to Enable or Disable Inter-Site Selective Fallover

To enable or disable the Resource Group Recovery with Selective Fallover behavior:

1. In SMIT, select **Extended Configuration > Extended Resource Configuration > Extended Resource Configuration > Customize Resource Group and Resource Recovery > Customize Inter-site Resource Group Recovery** and press Enter.

A selector screen lists the resource groups that contain nodes from more than one site (including those with a site management policy of **Ignore**. These are not affected by this function even if you select one of them.)

2. Select the resource groups for recovery customization.

   The next screen lists the selected resource groups and includes the field to enable or disable inter-site selective fallover.

3. To enable inter-site selective fallover (initiated by the Cluster Manager), select **true.** The default is **false** for a cluster migrated from a previous release, and **true** for a new HACMP 5.3 or 5.4 cluster.

Reliable NFS Function

You can configure NFS in all non-concurrent resource groups. See the chapter on Planning Shared LVM Components in the *Planning Guide* for information on planning prerequisites for configuring NFS as resources in resource groups.

As you configured resources, you can specify the following items related to NFS:

- Use the Reliable NFS server capability that preserves locks and dupcache (two-node clusters only).
- Specify a network for NFS mounting.
- Define NFS exports and mounts at the directory level.
- Specify export options for NFS-exported directories and filesystems.

Relinquishing Control over NFS Filesystems in an HACMP Cluster

Once you configure resource groups that contain NFS filesystems, you relinquish control over NFS filesystems to HACMP.

Once NFS filesystems become part of resource groups that belong to an active HACMP cluster, HACMP takes care of cross-mounting and unmounting the filesystems, during cluster events, such as fallover of a resource group containing the filesystem to another node in the cluster.

If for some reason you stop the cluster services and must manage the NFS filesystems manually, the filesystems must be unmounted before you restart the cluster services. This enables management of NFS filesystems by HACMP once the nodes join the cluster.

NFS Exporting Filesystems and Directories

The process of NFS-exporting filesystems and directories in HACMP differs from that in AIX 5L. The sections in Chapter 6: Planning Shared LVM Components in the *Planning Guide* explain the NFS-exporting process in HACMP. The following sections provide additional information.

Specifying Filesystems and Directories to NFS Export

In AIX 5L, you list filesystems and directories to be NFS-exported in the **/etc/exports** file; in HACMP, you must put these in a resource group.

You can configure NFS in all non-concurrent resource groups. See Chapter 6: Planning Shared LVM Components in the *Planning Guide* for information on planning prerequisites for configuring NFS as resources in resource groups.

Specifying Export Options for NFS Exported Filesystems and Directories

If you want to specify special options for NFS-exporting in HACMP, you can create a **/usr/es/sbin/cluster/etc/exports** file. This file has the same format as the regular **/etc/exports** file used in AIX 5L.

Use of this alternate exports file is optional. HACMP checks the **/usr/es/sbin/cluster/etc/exports** file when NFS-exporting a filesystem or directory. If there is an entry for the filesystem or directory in this file, HACMP will use the options listed. If the filesystem or directory for NFS-export is not listed in the file, or, if the user has not created the **/usr/es/sbin/cluster/etc/exports** file, the filesystem or directory will be NFS-exported with the default option of root access for all cluster nodes.

Configuring the Optional /usr/es/sbin/cluster/etc/exports File

In this step, you add the directories of the shared filesystems to the exports file. Complete the following steps for each filesystem you want to add to the exports file. Refer to your NFS-Exported Filesystem Worksheet.

Remember that this alternate exports file does not specify *what* will be exported, only *how* it will be exported. To specify what to export, you must put it in a resource group.

To add a directory to Exports List:

1. In SMIT, enter the fastpath `smit mknfsexp`.
   The system displays the **Add a Directory to Exports List** panel.
2. In the **EXPORT directory now**, **system restart or both** field, enter **restart**.
3. In the **PATHNAME of alternate Exports file** field, enter `/usr/es/sbin/cluster/etc/exports`. This step creates the alternate exports file which will list the special NFS export options.
4. Add values for the other fields as appropriate for your site, and press Enter. Use this information to update the **/usr/es/sbin/cluster/etc/exports** file.
5. Return to the **Add a Directory to Exports List** panel, or exit SMIT if you are finished.
6. Repeat steps 1 through 4 for each filesystem or directory listed in the **FileSystems/Directories to Export** field on your planning worksheets.

Forcing a Varyon of Volume Groups

Forcing a varyon of volume groups is an option that you should use only with understanding of its consequences. This section describes the conditions under which you can safely attempt to forcefully bring a volume group online on the node, in the case when a normal varyon operation fails due to a loss of quorum.

For a complete overview of the forced varyon functionality *and* quorum issues, see the section Forcing a Varyon in Chapter 5: Planning Shared LVM Components in the *Planning Guide.*

We recommend to specify the **super strict** disk allocation policy for the logical volumes in volume groups for which forced varyon is specified. Configuring the **super strict** disk allocation policy for volume groups that may be forced on does the following:

- Guarantees that copies of a logical volume are always on separate disks

  and

- Increases the chances that forced varyon will be successful after a failure of one or more disks.

Note: You should apply the **super strict** disk allocation policy for disk enclosures in the cluster. You specify the **super strict** policy under the **Allocate each logical partition copy on a separate physical volume?** option in the **Add a Logical Volume**, or **Change/Show a Logical Volume** SMIT panels in AIX 5L. Also, if you are using the **super strict** disk allocation policy, specify the correct number of physical volumes for this logical volume and do not accept the default setting of 32 physical volumes.

Use independent disk enclosures that use logical volume mirroring; place logical volume mirror copies on separate disks that rely on separate power supplies, and use separate physical network interfaces to ensure access. This ensures that no disk is a single point of failure for your cluster.

You can specify a forced varyon attribute for:

- Volume groups on SSA or SCSI disks that use LVM mirroring where you want to NFS mount the filesystems
- Volume groups that are mirrored between separate RAID or ESS devices.

Note: Be aware that when the forced varyon facility is used successfully and the volume group is brought online on the node (using the one complete copy of the data that was found), the data that you recover by forcing a volume group to go online is guaranteed to be consistent, but not necessarily the latest.

Note: During runtime, for large volume groups (those with more than 256 disks), checking logical partition maps may take extra processing time. However, since this time delay occurs *only* when you select a forced varyon for a large volume group in the case when a normal varyon failed due to a lack of quorum, enduring a slow varyon process that enables data recovery is preferable to having no chance at all to activate the volume group.

When HACMP Attempts a Forced Varyon

For troubleshooting purposes, it is helpful to know under what conditions or cluster events HACMP attempts a forced varyon, when this is configured. In general, HACMP attempts a forced varyon in the event of a cluster failure. The following list contains examples of cluster event failures that can trigger a forced varyon:

- Cluster startup, normal varyon fails due to a loss of quorum on one of the disks.
- Nodes joining the cluster, normal varyon fails due to a loss of quorum on one of the disks.

- Node reintegration, normal varyon fails for concurrent resource groups.
- Selective fallover caused by an application or a node failure moves a resource group to a takeover node.
- Selective fallover caused by a loss of quorum for a volume group moves a resource group to a takeover node.

  When HACMP selectively moves a resource group for which a loss of quorum for a volume group error has occurred, it tries to bring the volume groups online on the takeover node. If a normal varyon process for volume groups fails at this point, and, if you have specified a forced varyon for the volume groups in this resource group, then, since quorum is lost, HACMP attempts a forced varyon operation.

  To summarize, for the cases where HACMP uses selective fallover to move the resource groups, the sequence of events would be the following:

  - If, after an **rg_move** event, a forced varyon is launched and is successful, the resource group remains online on the node to which it has been moved.
  - If, after an **rg_move** event, a forced varyon is launched and fails, selective fallover continues to move the resource group down the node chain.

  Note: If a resource failure occurs in a concurrent resource group, HACMP takes this resource group offline on a particular node. In this case, use the **clRGmove** utility to manually bring the resource group online on the node.

Avoiding a Partitioned Cluster

The forced option to activate a volume group must be used with care. Should the cluster become partitioned, each partition might force on the volume group and continue to run. In this case, two unequal copies of the data will be active at the same time. This situation can cause data divergence and does not allow a clean recovery. Were this to happen with a concurrent volume group, the consequences would be even worse, as the two sides of the cluster would have made uncoordinated updates.

To prevent cluster partitioning, configure multiple heartbeating paths. Where possible, use heartbeating through a disk path (TMSCSI, TMSSA or disk heartbeating).

Verification Checks for Forced Varyon

If you specified a forced varyon attribute for a resource group, and HACMP detects that the logical volumes are not being mirrored with the **super strict** disk allocation policy, HACMP a warns upon verification of cluster resources. In this case, a forced varyon operation may not succeed.

As part of the process, HACMP checks the logical partitions on each disk for each volume group:

- If it cannot find a complete copy of every logical volume for a volume group, an error message: "`Unable to vary on volume group <vg name> because logical volume <logical volume name> is incomplete`" displays in the **hacmp.out** file. In this case, a forced varyon operation fails and you will see an event error.

- If HACMP can find a complete copy for every logical volume for all volume groups in this resource group that require a forced varyon, it varies on the volume groups on the node in the cluster.

Testing Your Configuration

After you configure a cluster, you should test it before making it available in a production environment. For information about using Cluster Test Tool to test your cluster, see Chapter 8: Testing an HACMP Cluster.

Chapter 6: Configuring Cluster Events

The HACMP system is event-driven. An event is a change of status within a cluster. When the Cluster Manager detects a change in cluster status, it executes the designated script to handle the event and initiates any user-defined customized processing.

To configure cluster events, you indicate the script that handles the event and any additional processing that should accompany an event as described below. You can define multiple customized pre- and post-event scripts (for a particular cluster event). The environment variable EVENT_STAGE will be set to the appropriate value of *pre, post, notify,* or *recovery* when the corresponding event command is run.

The SMIT **HACMP Extended Event Configuration** menu includes:

- Considerations for Pre- and Post-Event Scripts
- Configuring Pre- and Post-Event Commands
- Configuring Pre- and Post- Event Processing
- Configuring User-Defined Events
- Tuning Event Duration Time Until Warning
- Configuring a Custom Remote Notification Method.

Considerations for Pre- and Post-Event Scripts

Take into account the following information when planning your pre- and post-event scripts.

Using Shell Environment Variables in Pre- and Post-Event Scripts

When writing your pre- or post-event script, none of the shell environment variables defined in **/etc/environment** will be available to your program. If you need to use any of these variables you must explicitly source them by including this line in your script:

```
. /etc/environment
```

event_error Now Indicates Failure on a Remote Node

In releases prior to HACMP 5.2, non-recoverable event script failures resulted in the **event_error** event being run on the cluster node where the failure occurred. The remaining cluster nodes did not indicate the failure.

With HACMP 5.2 and up, all cluster nodes run the **event_error** event if any node has a fatal error. All nodes log the error and call out the failing node name in the **hacmp.out** log file. If you have added pre- or post-event scripts for the **event_error** event, be aware that they are called on *each* node, not just on the failing node.

A new Korn shell environment variable that indicates the node where the event script failed, EVENT_FAILED_NODE, is set to the name of the node where the event occurred. Use this variable in your pre- or post-event scripts to locate the failure.

The variable LOCALNODENAME identifies the local node; if LOCALNODENAME is not the same as EVENT_FAILED_NODE, then the failure occurred on a remote node.

Parallel Processing of Resource Groups Affects Event Processing

When resource groups are processed in parallel, fewer cluster events occur in the cluster. In particular, only **node_up** and **node_down** events take place, and events such as **node_up_local**, or **get_disk_vg_fs** do not occur. (This is because HACMP uses other methods to process resources in parallel.) As a result, the use of parallel processing reduces the number of particular cluster events for which you can create customized pre- or post-event scripts. If you start using parallel processing for some of the resource groups in your configuration, be aware that your existing event scripts may not work for the resource groups. For more information, see Appendix B: Resource Group Behavior during Cluster Events in this Guide, and Chapter 7: Planning Events in the *Planning Guide*.

Dependent Resource Groups and the Use of Pre- and Post-Event Scripts

Prior to HACMP 5.2, to achieve resource group and application sequencing, system administrators had to build the application recovery logic in their pre- and post-event processing scripts. Every cluster would be configured with a pre-event script for all cluster events, and a post-event script for all cluster events.

Such scripts could become all-encompassing “case” statements. For instance, if you want to take an action for a specific event on a specific node, you need to edit that individual case, add the required code for pre- and post-event scripts, and also ensure that the scripts are the same across all nodes. (For instance, to ensure that all scripts are the same on all nodes, each script must contain the logic for all nodes and execute the “case” for the node on which it is being run.)

To summarize, even though the logic of such scripts captures the desired behavior of the cluster, they can be difficult to customize and even more difficult to maintain later on, when the cluster configuration changes.

If you are using pre-and post-event scripts or other methods, such as customized serial resource group processing to establish dependencies between applications that are supported by your cluster, then these methods may no longer be needed or can be significantly simplified. Instead, you can specify dependencies between resource groups in a cluster. For more information on how to configure resource group dependencies, see Configuring Dependencies between Resource Groups.

If you still want to customize behavior for some applications, consider adding a pre- or post-event script to the **resource_state_change** event. See Chapter 7: Planning Events in the *Planning Guide* for more details on this event.

Configuring Pre- and Post-Event Commands

To define your customized cluster event scripts:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Event Configuration** > **Configure Pre/ Post-Event Commands** > **Add a Custom Event Command** and press Enter.
3. Enter the field values as follows:

Cluster Event Name	Enter a name for the command. The name can have a maximum of 32 characters.
Cluster Event Description	Enter a short description of the event.
Cluster Event Script Filename	Enter the full pathname of the user-defined script to execute.

4. Press Enter to add the information to HACMPcustom class in the local HACMP Configuration Database (ODM).
5. Go back to the **Extended Configuration** menu and select **Extended Verification and Synchronization** to synchronize your changes across all cluster nodes.

   Note: Synchronizing does not propagate the actual new or changed scripts; you must add these to each node manually.

Configuring Pre- and Post- Event Processing

Complete the following steps to set up or change the processing for an event. In this step you indicate to the Cluster Manager to use your customized pre- or post-event commands. You only need to complete these steps on a single node. The HACMP software propagates the information to the other nodes when you verify and synchronize the nodes.

To configure pre- and post-events for customized event processing:

1. Enter `smit hacmp`
2. Select **HACMP Extended Configuration > Extended Event Configuration** > **Change/Show Pre-defined HACMP Events** to display a list of cluster events and subevents.
3. Select an event or subevent that you want to configure and press Enter. SMIT displays the panel with the event name, description, and default event command shown in their respective fields.
4. Enter field values as follows:

Event Name	The name of the cluster event to be customize.
Description	A brief description of the event's function. This information cannot be changed.

Event Command	The full pathname of the command that processes the event. The HACMP software provides a default script. If additional functionality is required, it is strongly recommended that you make changes by adding pre-or post-event processing of your own design, rather than by modifying the default scripts or writing new ones.
Notify Command	(*Optional*) Enter the full pathname of a user-supplied script to run before and after a cluster event. This script can notify the system administrator that an event is about to occur or has occurred. The arguments passed to the command are: the event name, one keyword (either *start* or *complete*), the exit status of the event (if the keyword was *complete*), and the same trailing arguments passed to the event command.
Pre-Event Command	(*Optional*) If you have defined custom cluster events, press F4 for the list. Or, enter the name of a custom-defined event to run before the HACMP cluster event command runs. This command is run before the “event command” script is run. The arguments passed to this command are the event name and the trailing arguments passed to the event command. Remember that the Cluster Manager will not process the event until this pre-event script or command has completed.
Post-Event Command	(*Optional*) If you have defined custom cluster events, press F4 for the list. Or, enter the name of the custom event to run after the HACMP cluster event command executes successfully. This script provides post-processing after a cluster event. The arguments passed to this command are the event name, event exit status, and the trailing arguments passed to the event command.
Recovery Command	(*Optional*) Enter the full pathname of a user-supplied script or AIX 5L command to execute to attempt to recover from a cluster event command if you are using Switched Fabric. If the recovery command succeeds and the retry count is greater than zero, the cluster event command is rerun. The arguments passed to this command are the event name and the arguments passed to the event command.

Recovery Counter	Enter the number of times to run the recovery command. Set this field to zero (0) if no recovery command is specified, or to at least one (1) if a recovery command is specified.

5. Press Enter to add this information to the HACMP Configuration Database.
6. Return to the **HACMP Extended Configuration** panel and synchronize your event customization by selecting the **Extended Verification and Synchronization** option. Note that all HACMP event scripts are maintained in the **/usr/es/sbin/cluster/events** directory. The parameters passed to a script are listed in the script's header.

   Note: You or a third-party system administrator can reset the HACMP tunable values, such as cluster event customizations, to their installation-time defaults. For more information, see the Resetting HACMP Tunable Values section in Chapter 1: Troubleshooting HACMP Clusters in the *Troubleshooting Guide*.

See Chapter 10: Monitoring an HACMP Cluster, for a discussion on how to emulate HACMP event scripts without actually affecting the cluster.

Configuring User-Defined Events

Note: Changes to custom user-defined events are not supported in an active cluster. User-defined events are not supported by a dynamic reconfiguration of the cluster. You must manually distribute the HACMPude ODM to all nodes after you make any changes.

To add a user-defined event:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Event Configuration > Configure User-Defined Events** > **Add Custom User Defined Event** panel.
3. Enter the field values as follows:

Event name	The name of the cluster event.
Recovery program path	The full pathname of the recovery program.
Resource attribute	A string consisting of a resource class name followed by a period followed by a resource attribute.
Select String	A set of elements whose values identify the scope of the resource attribute.
Expression	The relational expression between resource persistent attributes and other elements (such as constants) that, when true, generates an event.

Rearm expression	An expression used to generate an event that alternates with an original event expression in the following way: The event expression is used until it is true, then the rearm expression is used until it is true, then the event expression is used, and so on. The rearm expression is commonly the inverse of the event expression (for example, a resource variable is on or off). It can also be used with the event expression to define an upper and lower boundary for a condition of interest.

Changing or Showing User-Defined Events

To verify that the existing event definitions are specified as intended after a migration from a previous release of HACMP, use the SMIT panel. Do *not* use the **odmget** command for this purpose. The **odmget** command displays strings for the event definitions that differ from the information that was entered in SMIT. This is the expected behavior.

When defining the selection string in SMIT, you don't need the escape characters for this expression. HACMP will add those necessary to get the event into and out of the ODM. Specify the event as in the following example:

* Event name	fsfull
* Recovery program path	[/tmp/fsfullevent.rp]
* Resource name	[IBM.FileSystem]
* Selection string	[Name == “/tmp”]
* Expression	[PercentTotUsed>65]
Rearm expression	[PercentTotUsed<65]

Note for this event if you issue the **odmget HACMPude** command, the output displays the escape characters:

```
HACMPude:
        name = "fsfull"
        state = 0
        recovery_prog_path = "/tmp/fsfullevent.rp"
        recovery_type = 2
        recovery_level = 0
        res_var_name = "IBM.FileSystem"
        instance_vector = "Name == \"/tmp\""
        predicate = "PercentTotUsed>65"
        rearm_predicate = "PercentTotUsed<65"
```

To change a custom user-defined event, or to show a list of the events currently defined:

1. From the **Configure User Defined Event** panel, select **Change/Show Custom User Defined Event**. SMIT displays a list of all currently defined custom events.
2. Select the event to change or view and press Enter.

SMIT displays the **Change/Show Custom User Defined Event** panel, with the currently defined information about the event.

3. Change any information, and then press Enter.

Removing User-Defined Events

To remove a custom user-defined event:

1. From the **Configure User Defined Event** panel, select **Remove Custom User Defined Event**. SMIT lists all currently defined custom events.
2. To remove a particular event, select the event and press Enter.

   SMIT displays the currently defined information about the event.
3. Press Enter to remove the event.

Tuning Event Duration Time Until Warning

Depending on cluster configuration, the speed of cluster nodes and the number and types of resources that need to move during cluster events, certain events may take different times to complete. Cluster events run asynchronously and usually call AIX 5Lsystem commands. Since HACMP has no means to detect whether the event script is actually performing useful work at a given period of time, it runs a **config_too_long** event (which sends messages to the console and to the **hacmp.out** file) each time the processing of the event exceeds a certain amount of time. For such events, you may want to customize the time period HACMP waits for an event to complete before issuing the **config_too_long** warning message.

Also, see the section on this topic in Chapter 7: Planning for Cluster Events in the *Planning Guide* for more information on when to alter the time before receiving a system warning.

Note: The **config_too_long** warning timer for **node_up** should be adjusted to allow for longer time to process **node_up** events with dependent resource groups. **node_up** processing in clusters with dependencies could take more time than in the clusters without dependent resource groups.

Prerequisites and Notes

The following are important to keep in mind when you are working with event duration:

- The total duration time is calculated differently for “slow” and “fast” cluster events.

  “Fast” events are those that do not include acquiring or releasing resources and normally take a shorter time to complete.

  For “fast” events, the total duration time during which HACMP waits before issuing a warning is equal to Event Duration Time.

  “Slow” cluster events are those that involve acquiring and releasing resources, use application server start and stop scripts, or site events using HAGEO. “Slow” events may take a longer time to complete. Customizing event duration time for “slow” events lets you avoid getting unnecessary system warnings during normal cluster operation.

For “slow” events, the total duration time before receiving a **config_too_long** warning message is set to the sum of **Event-only Duration Time** and **Resource Group Processing Time**.

- Remember, you can customize event duration time before receiving a warning for cluster events, *not* for nodes or specific resource groups in your cluster. Once the **Total Event Duration Time** is specified, the system waits for the specified period of time and sends a **config_too_long** message to the node which was affected by this event.

  For example, you have a cluster with five resource groups. A **node_down** event (a “slow” event) occurs on Node A, which owns some of the resource groups. And, you have previously specified the **Event-only Duration Time** to be 120 seconds, and the **Resource Group Processing Time** to be 400 seconds.

  When a **node_down** event occurs on Node A, a **config_too_long** message is sent to Node A according to this formula:

  `Event Duration Time (120 seconds) + Resource Group Processing Time (400 seconds) = 520 seconds (Total Event Duration Time).` A **config_too_long** message appears on Node A after 520 seconds.
- During dynamic reconfiguration events, the Cluster Manager uses the previously specified values of the event duration time until warning. After dynamic reconfiguration is complete and the new values of event duration time get synchronized, the Cluster Manager uses the newly specified values.

You can configure **Event Duration Time** using the **HACMP for AIX > Extended Configuration > Extended Event Configuration > Change/Show Time Until Warning** panel in SMIT.

Changing Event Duration Time Until Warning

To change the total event duration time before receiving a **config_too_long** warning message, perform the following procedure on any cluster node:

1. Enter `smit hacmp`
2. In SMIT, select **HACMP Extended Configuration > Extended Event Configuration > Change/Show Time Until Warning** and press Enter.
3. Enter data in the fields as follows:

Max. Event-only Duration (in seconds)	Enter any positive integer. This is the maximum time (in seconds) it takes to execute a cluster event. The default **Max. Event-only Duration** is 180 seconds. For “fast” cluster events, such as events that do not involve acquiring or releasing resource groups, the total event duration time before HACMP issues a warning is equal to the **Max. Event-only Duration.**

Max. Resource Group Processing Time (in seconds)	Enter any positive integer or zero. This is the maximum time (in seconds) it takes to acquire or release a resource group. The default **Max. Resource Group Processing Time** time is 180 seconds. Note that if you have several resource groups that have to be acquired or released during a cluster event, the value in this field should be longer than the maximum acquisition or release time for any of the cluster resource groups. For "slow" cluster events, such as events that include acquiring or releasing resources, the total event duration time (before HACMP issues a warning) is equal to the *sum* of **Max**. **Resource Group Processing Time** and **Max. Event-only Duration**.
Total time to process a Resource Group Event before a warning is displayed	The total time for the Cluster Manager to wait before running the **config_too_long** script. The default is 6 minutes and 0 seconds. This field is the sum of the two other fields and is not editable.

4. Press Enter to change the field values. HACMP changes these values in the HACMP Configuration Database.
5. Synchronize the cluster to propagate the data to the other cluster nodes. HACMP uses the specified total event duration time before issuing **config_too_long** warning messages.

Configuring a Custom Remote Notification Method

You can configure a remote notification method through SMIT to issue a customized numeric or alphanumeric page in response to a specified cluster event. Starting with HACMP 5.3, you can also send SMS text message notifications to any address, including a cell phone SMS address or mail to an email address. The pager message is sent through the attached dialer modem. Cell phone text messages are sent through email using the TCP/IP connection or an attached GSM wireless modem.

The following sections describe how to configure custom remote notification methods to respond to an event, how cluster verification confirms the remote notification configuration**,** and how node failure affects the remote notification method.

You can send the following custom remote notifications:

- Numeric and alphanumeric page
- SMS text message to any address including a cell phone or mail to an email address.
- SMS text message using a GSM modem to transmit the notification through a wireless connection.

Prerequisites

The HACMP remote notification functionality requirements follow:

- A tty port used for paging cannot also be used for heartbeat traffic or for the DBFS function of HAGEO.

- Any tty port specified must be defined to AIX 5L and must be available.
- Each node that may send a page or text messages must have an appropriate modem installed and enabled.

  Note: HACMP checks the availability of the tty port when the notification method is configured and before a page is issued. Modem status is *not* checked.

  Note: To send an SMS text message over the dialer modem, your pager provider must offer this service.

- Each node that may send email messages from the SMIT panel using AIX 5L mail must have a TCP/IP connection to the Internet.
- Each node that may send text messages to a cell phone must have an appropriate Hayes-compatible dialer modem installed and enabled.
- Each node that may transmit an SMS message wirelessly must have a Falcom-compatible GSM modem installed in the RS232 port with the password disabled. Ensure that the modem connects to the cell phone system.

Creating a Remote Notification Message File

Before you can issue a message to a pager or cell phone, you must create a file that contains the message text. HACMP provides a template to help you create this file. The template contains default text and instructions for an alphanumeric page or cell phone message. The template is in:

```
/usr/es/sbin/cluster/samples/pager/sample.txt
```

By default, the message contains the following information: the event, the node on which it occurred, the time and date, and the name of the object (node, network, site, etc.) affected by the event. This default message is sent if no message file is found at the time a custom alphanumeric page or cell phone message is triggered.

For numeric pages, the provided sample text is not appropriate; your numeric page file should contain only digits. If no message file is found when a numeric page is triggered, the default message sent is “888.”

The **sample.txt** file contains comments that relate to an alphanumeric pager or cell phone message. A numeric page does not use this file. Shown below is the **sample.txt** file; there is no need to alter the file unless you want to add additional recipients.

Note: Save the **sample.txt** file with a new name before modifying it. However, if you do alter the file when you migrate to a new version of HACMP, the customized file is preserved, even though a new default **sample.txt** file is installed. See the related section in the *Installation Guide* on upgrading to HACMP 5.4 for details on where your modified **sample.txt** file is saved after a new installation.

Note: Place a separate copy of each message file on each node listed in the notification method definition. HACMP does *not* automatically distribute this file to other nodes during cluster synchronization.

Contents of the Sample.txt file

The following lists the contents of the sample.txt file:

```
# sample file for alphanumeric paging
# you can use the following notations in your message
# %d - current time&date
# %n - node that sends the message
# %e - eventname
# '#' is used to comment the line
# for example "Node %n: Event %e occured at %d"
# if nodename=bazilio, event=node_up
# and current date&time = Thu Sep 28 19:41:25 CDT 2006
# will result in sending the message
# "Node bazilio: Event node_up occured at Thu Sep 28 19:41:25 CDT 2006"
```

Defining a Remote Notification Method

To define a pager notification method first define a tty port for each node that might issue the page, and then define the remote notification method. To define a cell phone text message, follow the steps listed in the section Defining a New Remote Notification Method.

Defining a TTY Port to Issue a Page

To define a tty port for each node that might issue a page:

1. Enter `smit hacmp`
2. In SMIT, select **HACMP Extended Configuration > HACMP Extended Event Configuration > Configure Remote Notification Method** and press Enter.

   (You can also reach this panel by typing `smit cl_pager`.)
3. Select **Configure Node/Port Pairs**.
4. Select the node that will issue the page from the list of cluster nodes.
5. Press F4 for a list of available ports for the chosen node, and select *one* port from the list.
6. Repeat steps 3 and 4 for each node that might be called to issue a page.

Defining a New Remote Notification Method

To define a new remote notification method:

1. n SMIT, select **HACMP Extended Configuration > HACMP Extended Event Configuration > Configure Remote Notification Method > Add a Custom Remote Notification Method** and press Enter.
2. Fill in field values as follows:

Method Name	Assign a name to the notification method. This could also indicate who would get the message.
Description	Add a description, if desired, of the notification method.

Nodename(s)

Enter the name(s) of one or more nodes that you want to issue this or cell phone message. Press F4 to get a list of node names. Each node must have been defined previously in the **Define Port/Node Pairs** SMIT panel. Separate multiple nodes with a space.

Note: The sequence of nodes in this SMIT field determines their priority for sending pages or cell phone messages.

See Remote Notification and Node Failure for more information on node priority for remote notification.

Number to Dial or Cell Phone Address

Indicate the telephone number to dial to reach the pager or the address of the cell phone. The number-to-dial string can contain any characters or sequences supported by a Hayes-compatible modem using the standard Telocator Alphanumeric Protocol (TAP)—your provider must support this service.

- Depending on the type of pager, you will need to enter either the number of the pager alone, or the number of the paging company followed by the pager number:

  If you are using a numeric pager, use the form:
  `18007650102,,,,`
  The commas create pauses in the dialing sequence. The trailing commas are required because there is always some delay between dialing and the actual sending of the page.

  If the pager is alphanumeric the input should take the form:
  `180007654321;2119999`
  where 18007654321 is the paging company number and 2119999 is the actual pager number.
- For cell phone text messaging using email, enter the address of the cell phone. This is in the format: `phone_number@provider_address`. Consult your provider for the specific provider_address format. It may look like `180007654321@provider.net`. Multiple space-separated addresses can be used. Test this by sending an email. To send email to multiple addresses, separate the addresses using a space.
- You can send a text message wirelessly to a cell phone, if a GSM modem is used instead of the dialer modem. The format is <cell phone number>#. For example, it may look like 7564321#.

  The SIM providers may support international calls.

Filename

Specify the path of the text file containing the pager message or cell phone message.

Note: Make sure the path refers to the correct location of the message file on each node specified in the **Node Name(s)** field.

Cluster Event(s)	Specify the event(s) that activate this notification method. Press F4 to get a list of event names. Separate multiple events with a space.
Retry Counter	Specify how many times to reissue the page or cell phone message if it fails. The default is **3** times.
TIMEOUT	Specify how many seconds to wait before considering a page or cell phone message attempt failed. The default is **45** seconds.

3. When you finish entering values in all fields, press Enter.
4. Synchronize the cluster to propagate the configuration information to the other nodes.

Note: The configuration information can be entered on one node and propagated to the others during synchronization, but you must manually make sure that the correct page or cell phone message text files reside in the correct locations on each node in the nodelist.

Verification of Remote Notification Methods

When you synchronize or perform cluster verification, HACMP checks the configuration of your remote notification method and issues an error in these cases:

- A specified pager or cell phone message file is missing from a node it should reside on. (The message can still be sent—it will contain the text supplied in the original **sample.txt** file.)
- The same tty is defined for both heartbeat traffic and paging.
- The same tty is defined for both DBFS and paging.

Sending a Test Remote Notification Message

You can send a test page or cell phone message to make sure everything is configured correctly, and that the expected notification will be issued for a given event, just as if the event actually occurred.

Before sending the test remote message, you must have a notification method already configured. The test remote message must be sent from a node that is configured for the selected method.

To configure a remote notification message:

1. From the **Configure Custom Remote Notification Method** menu, select **Send a Test Remote Message.**
2. Select a remote notification method to use for the test.
3. In the **Send a Test Remote Message** panel, fill in field values as follows:

Method Name	The configured method that you selected for the test page.
Event name	Press F4 to get a picklist of events configured for the selected method, and select the event for which you would like to send a test page.

4. Press Enter. The Command Status window then reports the remote message was successful, or errors occurred.

The test remote message will be the message file you specified when you configured the notification method. If an object name is included, for the test remote message, it will appear as a pseudo-name such as node_1, adapter_1, site_1, network_1, etc. If the message file cannot be located, a default message will be sent to the pager or cell phone and an error will be displayed. For alphanumeric pages or cell phone messages, the default message is the sample text; for numeric pages, the default message is "888."

Remote Notification and Node Failure

If a node fails and triggers a page or cell phone message, the remote notification is sent from the node with the next highest priority. (A node's order of priority is determined by the order in which you listed node names when you defined the method.) If the next highest priority node is up but unable to send the remote notification for some other reason (for instance, the modem is not connected), the system attempts to resend the remote notification message for the number of times specified in the **Retry Counter**. If the remote notification still cannot be sent, it fails. The remote notification is *not* passed on to be issued from another node.

Changing or Removing a Custom Remote Notification Method

You can change or remove a notification method through SMIT to issue a customized remote notification in response to a specified cluster event.

Changing a Remote Notification Method

To change the configuration of a custom remote notification method:

1. Enter `smit hacmp`
2. In SMIT, select **HACMP Extended Configuration > Extended HACMP Event Configuration> Configure Remote Notification Methods > Change/Show Custom Remote Notification Method** and press Enter.

   (You can also reach the **Configure Remote Notification Methods** panel by typing `smit cl_pager`.)
3. Select the method you want to change.
4. Make your changes.
5. Press Enter.

Deleting a Remote Notification Method

To delete a custom remote notification method:

1. Enter `smit hacmp`
2. In SMIT, select **HACMP Extended Configuration > Extended HACMP Event Configuration> Configure Remote Notification Methods > Remove Custom Remote Notification Method** and press Enter.
3. Specify the name of the method you want to delete.
4. Press Enter to delete the method.

Chapter 7: Verifying and Synchronizing an HACMP Cluster

Verifying and synchronizing your HACMP cluster assures you that all resources used by HACMP are configured appropriately and that rules regarding resource ownership and resource takeover are in agreement across all nodes. You should verify and synchronize your cluster configuration after making *any* change within a cluster (for example, any change to the hardware operating system, node configuration, or cluster configuration).

The main sections of this chapter include the following:

- Overview
- Automatic Verification and Synchronization
- Verifying the HACMP Configuration Using SMIT
- Managing HACMP File Collections
- Adding a Custom Verification Method
- List of Reserved Words.

Overview

Whenever you configure, reconfigure, or update a cluster, run the cluster verification procedure to ensure that all nodes agree on the cluster topology, network configuration, and the ownership and takeover of HACMP resources. If the verification succeeds, the configuration can be synchronized. Synchronization takes effect immediately on an active cluster. A dynamic reconfiguration event is run and the changes are committed to the active cluster.

Note: If you are using the SMIT **Initialization and Standard Configuration** path, synchronization automatically follows a successful verification. If you are using the **Extended Configuration** path, you have more options for types of verification. If you are using the **Problem Determination Tools** path, you can choose whether to synchronize or *not*.

The messages output from verification indicate where the error occurred (for example, the node, device, or command). The utility uses verbose logging to write to the **/var/hacmp/clverify/clverify.log** file.

Note: Verification is *not* supported on a mixed-version HACMP cluster.

HACMP 5.4 has an additional verification check to ensure that each node can reach each other node in the cluster through non-IP network connections. If this is *not* true, a message is displayed.

Error conditions result when information is not properly configured on all cluster nodes. This information may be important for the operation of HACMP, but *not* part of the HACMP software itself; for example, AIX 5L volumes do *not* exist in the cluster. In some of these situations, you can authorize a corrective action before verification continues. When verification detects certain conditions, such as mismatched HACMP shared volume group time stamps or a node is missing required entries in **/etc/services**, HACMP fixes the problem. For a list of all conditions for which HACMP issues automatic corrective actions, see Running Corrective Actions during Verification.

On the node where you run the utility, detailed information is collected into log files, which contain a record of all data collected and the tasks performed.

You can add your own custom verification methods to ensure that specific components within your cluster are properly configured. You can change or remove these methods from the verification process depending on the level of cluster verification you want. See the section Adding a Custom Verification Method later in this chapter.

Note: Verification requires 4 MB of disk space in the **/var** filesystem in order to run; 18 MB of disk space is recommended for a four-node cluster. Typically, the **/var/hacmp/clverify/clverify.log** files require 1–2 MB of disk space.

Running Cluster Verification

After making a change to the cluster, you can perform cluster verification in the following ways:

- *Automatic verification.* You can automatically verify your cluster:
  - Each time you start cluster services on a node
  - Each time a node rejoins the cluster
  - Every 24 hours.

  By default, automatic verification is enabled to run at midnight.

  For detailed instructions, see Automatic Verification and Synchronization.
- *Manual verification.* Using the SMIT interface, you can either verify the *complete* configuration, or only the *changes* made since the last time the utility was run.

  Typically, you should run verification whenever you add or change anything in your cluster configuration. For detailed instructions, see Verifying the HACMP Configuration Using SMIT.

Automatic Verification and Synchronization

During *automatic verification and synchronization*, HACMP discovers and corrects several common configuration issues prior to starting cluster services. This automatic behavior ensures that if you had *not* manually verified and synchronized your cluster prior to starting cluster services, HACMP will do so. Throughout this section, automatic verification and synchronization is often simply referred to as *verification*.

Understanding the HACMP Cluster Verification Process

By default, verification runs automatically without any configuration required. We recommend that you do *not* disable verification, but if necessary, you can disable it using the **Extended Configuration >Extended Cluster Service Settings** path.

Verification occurs on both active and inactive clusters. In order for automatic verification to work, more than one node must exist in the cluster, since HACMP compares the configuration of one node against the configuration of another node.

Verification ensures an error-free cluster startup and poses a negligible impact on performance, which is directly related to the number of nodes, volume groups, and filesystems in the cluster configuration.

The phases of the verification and synchronization process are as follows:

1. Verification
2. Snapshot (optional)
3. Synchronization.

For details on these phases, see the Understanding the Detailed Phases of Verification section. After verification, cluster services start.

Cluster Verification during a Dynamic Cluster Reconfiguration Event

If a node is down during a dynamic reconfiguration event and later it attempts to join the cluster, cluster verification and synchronization run prior to starting services on the joining node, and the joining node receives its configuration updates from an active cluster node.

If verification fails on the joining node, the node will *not* start cluster services. Likewise, if a node is dynamically removed from the active cluster, the node will *not* be allowed to join the cluster or cluster services.

Parameters Automatically Corrected

Automatic verification and synchronization ensure that typical configuration inconsistencies are automatically corrected as follows:

- RSCT versions are congruent across the cluster.
- IP addresses (that RSCT expects) are configured on the network interfaces.
- Shared volume groups are *not* set to automatically varyon.
- Filesystems are *not* set to automatically mount.

Verifying RSCT Versions

The activity state of your nodes determines which RSCT number is used for synchronization. The number from the active nodes is used to populate the inactive nodes; if cluster services are currently running, it is assumed that all RSCT numbers are correct, so they are *not* verified.

If there are no active nodes and the number is inconsistent across the cluster, then verification uses the local node RSCT number to synchronize to all other cluster nodes—except if the local node RSCT number is zero (0), then HACMP uses 1 on all other cluster nodes.

Verifying Service IP Address Aliases

At cluster startup, RSCT expects the IP address label to be defined on the interfaces with the same value that has been defined in the HACMP configuration database. The HACMP automatic verification and synchronization process ensures nodes *not* currently running cluster services are verified and corrected; nodes currently running cluster services are *not* automatically corrected.

Note: Only aliased IP interfaces that are used by HACMP are verified and corrected.

If a node has an interface that is *not* defined as it appears in the HACMP configuration database, automatic verification detects this and issues an error message.

Verifying Shared Volume Groups

Shared volume groups that are configured as part of an HACMP resource group must have their automatic varyon attribute set to **No**. If the verification phase determines that the automatic varyon attribute is set to **Yes**, verification notifies you about nodes on which the error occurs and prompts you to correct the situation.

Verifying Filesystems

Any filesystems participating in a resource group with AIX 5L attributes that allow the filesystem to be automatically mounted at system restart will raise errors. This includes standard journaled filesystems (JFS) and enhanced journaled filesystems (JFS2). If the filesystem has been set to mount automatically at boot time, verification displays an error.

Understanding the Detailed Phases of Verification

This section describes the phases of verification and cluster services startup. These events occur in the following order:

- Phase One: Verification
- Phase Two: (Optional) Snapshot
- Phase Three: Synchronization.

After verification, cluster services start up. If cluster services do *not* start, it is because HACMP has discovered errors. You can resolve these errors by correcting inconsistencies. For information about correcting these inconsistencies, see the section Monitoring Verification and Resolving Configuration Inconsistencies in this chapter.

Phase One: Verification

During the verification process the default system configuration directory (DCD) is compared with the active configuration. On an inactive cluster node, the verification process compares the local DCD across all nodes. On an active cluster node, verification propagates a copy of the active configuration to the joining nodes.

If a node that was once previously synchronized has a DCD that does *not* match the ACD of an already active cluster node, the ACD of an active node is propagated to the joining node. This new information does *not* replace the DCD of the joining nodes; it is stored in a temporary directory for the purpose of running verification against it.

Note: When you attempt to start a node that has an invalid cluster configuration, HACMP transfers a valid configuration database data structure to it, which may consume 1–2 MB of disk space.

If the verification phase fails, cluster services will *not* start. In this situation, see the section Monitoring Verification and Resolving Configuration Inconsistencies.

Phase Two: (Optional) Snapshot

A snapshot is only taken if a node request to start requires an updated configuration. During the snapshot phase of verification, HACMP records the current cluster configuration to a snapshot file—for backup purposes. HACMP names this snapshot file according to the date of the snapshot and the name of the cluster. Only one snapshot is created per day. If a snapshot file exists and its filename contains the current date, it will *not* be overwritten.

This snapshot is written to the **/usr/es/sbin/cluster/snapshots/** directory.

The snapshot filename uses the syntax **MM-DD-YYYY-*ClusterName*-autosnap.odm**. For example, a snapshot taken on April 2, 2006 on a cluster hacluster01 would be named **usr/es/sbin/cluster/snapshots/04-02-06hacluster01-autosnap.odm**.

Phase Three: Synchronization

During the synchronization phase of verification, HACMP propagates information to all cluster nodes. For an inactive cluster node, the DCD is propagated to the DCD of the other nodes. For an active cluster node, the ACD is propagated to the DCD.

If the process succeeds, all nodes are synchronized and cluster services start. If synchronization fails, cluster services do *not* start and HACMP issues an error.

Monitoring Verification and Resolving Configuration Inconsistencies

You can monitor the automatic verification and synchronization progress as it occurs by tracking messages as they appear on the SMIT console. In addition, you can examine any prior processes by reviewing the **smit.log** file or **/var/hacmp/clverify/clverify/log**.

Verification Completion

When cluster verification completes on the selected cluster node, this node supplies the following information to the other cluster nodes:

- Name of the node where verification had been run
- Date and time of the last verification
- Results of the verification.

This information is stored on every available cluster node in the **/var/hacmp/clverify/clverify.log** file. If the selected node became unavailable or could *not* complete cluster verification, you can detect this by the lack of a report in the /**var/hacmp/clverify/clverify.log** file. If the log file does *not* indicate a specific node, then the error applies to all nodes and cluster services do *not* start.

If cluster verification completes and detects some configuration errors, you are notified about the potential problems:

- The exit status of verification is published across the cluster along with the information about cluster verification process completion.
- Broadcast messages are sent across the cluster and displayed on the console. These messages inform you of any detected configuration errors.
- A **cluster_notify** event runs on the cluster and is logged in **hacmp.out** (if cluster services are running).
- Information about the node where you ran the cluster verification is written to the **/var/hacmp/clverify/clverify.log** file. If a failure occurs during processing, error messages and warnings indicate the node affected and reasons for the verification failure.
- A configuration snapshot is written to the **/usr/es/sbin/cluster/snapshots/** directory.

Ongoing Automatic Verification

Once a valid configuration is defined, the verification process runs once every 24 hours. By default, the first node in alphabetical order runs the verification at midnight; however, you can change these defaults by selecting a node and a time that suits your needs. If the selected node is unavailable (powered off), automatic verification does *not* run.

For information on changing the default configuration see the Automatic Cluster Configuration Monitoring section in Chapter 1: Troubleshooting HACMP Clusters in the *Troubleshooting Guide*.

Verifying the HACMP Configuration Using SMIT

After reconfiguring or updating a cluster, run the cluster verification procedure. For a list of the types of verification performed, see Verifying and Synchronizing a Cluster Configuration.

Note: If you are investigating a problem with the cluster and want to run verification procedures without synchronizing the cluster, use the cluster verification SMIT panels found under the **Problem Determination Tools** menu. See Chapter 1: Troubleshooting HACMP Clusters in the *Troubleshooting Guide*.

Verifying and Synchronizing a Cluster Configuration

Verification performs many automatic checks. This section provides overviews of the following verifications performed; it is *not* an exhaustive description of all verifications. HACMP documentation lists the verification checks for each function in the sections describing these functions.

This section includes the following topics:

- Verifying the Topology Configuration
- Verifying the Network Configuration
- Verifying Disk and Filesystem Configuration
- Verifying Resource Group Information
- Verifying Individual Resources
- Verifying Automatic Error Notification Methods
- Verifying the Security Configuration
- Verifying Custom Configurations
- Verifying HACMP/XD Configurations
- Verifying Service IP labels.

Verifying the Topology Configuration

Verification ensures that all nodes agree on the topology of the cluster. For example, it checks for invalid characters in cluster names, node names, network names, network interface names, and resource group names. It checks to ensure that interfaces are properly configured, nodes are reachable, and networks have the required number of interfaces.

It also checks for the reserved words used as cluster names, node names, network names, network interface names and resource group names. These names are listed in the **/usr/es/sbin/cluster/etc/reserved_words** file. See the List of Reserved Words in this chapter.

Verifying the Network Configuration

Verification ensures that the networks are configured correctly and that all nodes agree on the ownership of all defined resources, such as the following:

- Configuration of network information, such as addresses on all nodes in the cluster or whether multiple non-IP networks exist on the same tty device.
- No network interfaces configured on unsupported network types (for example, IP, socc, slip and fcs).

Verifying Disk and Filesystem Configuration

Verification ensures that disks and filesystems are in agreement and configured according to the following:

- Agreement among all nodes on the ownership of defined resources (for example, filesystems, volume groups, disks, and application servers). The verification utility checks for the existence and defined ownership of the filesystems to be taken over, and then checks the volume group and disks where the filesystems reside.

- Agreement among nodes on the major and minor device numbers for NFS-exported filesystems.
- If disk fencing is enabled, verification sends an error if all nodes are *not* included in the concurrent access resource group.

Verifying Resource Group Information

Verification ensures that the resource group information supplied is in agreement and configured according to the following:

- Verification issues warnings in cases when the startup, fallover or fallback preferences that you choose for resource groups may put the high availability of resources at risk in the case of a cluster failure.
- The verification utility checks that the choices for distribution of resources in case of a takeover (node priorities) so that the takeover information matches the owned resources information.

Verifying Individual Resources

Verification checks individual resources, such as the following:

- Event customization.
- Application server start and stop scripts exists and that they are executable.

Verifying Automatic Error Notification Methods

Verification ensures that automatic error notification (AEN) methods exist and are properly configured for the following:

- Root volume groups
- HACMP-defined volume groups or HACMP-defined disks
- HACMP-defined filesystems (the underlying disks that support the file system)
- SP switch network interface cards.

Verifying the Security Configuration

If you have configured Kerberos on your system, verification also verifies that:

- Kerberos is installed on all nodes in the cluster
- All IP labels listed in the configuration have the appropriate service principals in the **.klogin** file on each node in the cluster
- All nodes have the proper service principals
- All nodes have the same security mode setting.

Verifying Custom Configurations

If you have configured custom cluster snapshot methods or AIX 5L Fast Connect services, verification checks their existence and consistency.

Verifying HACMP/XD Configurations

If you are using HACMP/XD configurations, verification confirms that the HACMP/XD cluster configuration for sites and its replicated resources are consistent with your HACMP cluster configuration.

Verifying Service IP labels

If a service IP label is configured on the interface instead of the boot label, verification issues an error reminding you to run the sample utility **clchipdev** before starting cluster services. If that service IP label is an alias, verification has a correct action to reverse it. The sample utility **clchipdev** helps configure the application service interface correctly in HACMP. For information on **clchipdev,** see Configuring an Application Service Interface in Chapter 4: Configuring HACMP Cluster Topology and Resources (Extended).

Verifying and Synchronizing the Cluster Configuration

You can verify and synchronize your cluster from either SMIT cluster configuration path:

- Initialization and Standard Configuration
- Extended Configuration.

Verifying the Cluster Using the Initialization and Standard Configuration Path

If you use the **Initialization and Standard Configuration** path, when you select the option **Verify and Synchronize HACMP Configuration**, the command executes immediately. Messages appear in the SMIT command status screen as the configuration is checked.

Verifying the Cluster Using the Extended Configuration Path

If you use the **Extended Configuration** path, you can set parameters for the command before it runs. These parameters differ depending on whether or *not* the cluster is active.

To verify and synchronize the HACMP cluster configuration:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Verification and Synchronization** and press Enter.

   The software checks whether cluster services are running on any cluster node and displays one of the following screens:

   If the cluster is active, the following options appear.

Emulate or Actual	**Actual** is the default.
Verify changes only?	**No** is the default. (Run the full check on resource and topology configuration.) Select **Yes** to verify only resource or topology configurations that have changed since the last time the cluster was verified. **Note**: If you have changed the AIX 5L configuration, do *not* use this mode; it only applies to HACMP configuration changes.

Logging	**Standard** is the default. You can also select **Verbose**. Verification messages are logged to **/var/hacmp/clverify/clverify.log**.

If the cluster is inactive, the following options appear:

Verify Synchronize or Both	**Both** is the default. You can also select **Verify** only or **Synchronize** only.
Automatically correct errors found during verification?	**No** is the default. HACMP will *not* perform corrective actions. If you select **Interactively**, during verification you will be prompted when it finds a problem it can correct, for example: • Importing a volume group • Re-importing shared volume groups (mount points and filesystems issues). You then choose to have the action taken or *not*. For more information, see the section Conditions That Can Trigger a Corrective Action in this chapter.
Force synchronization if verification fails?	**No** is the default. If you select **Yes**, cluster verification runs but verification errors are ignored and the cluster is synchronized. Use the **Yes** option with caution. Correct functioning of the cluster at runtime can*not* be guaranteed if you synchronize without verification. Cluster topology errors may lead to an abnormal exit of the Cluster Manager. Resource configuration errors may lead to resource group acquisition errors.
Verify changes only?	**No** is the default. (Run the full check on resource and topology configuration.) **Yes** opts to verify only resource or topology configurations that have changed since the last time the cluster was verified. **Note**: If you have changed the AIX 5L configuration, do *not* use this mode; it only applies to HACMP configuration changes.
Logging	**Standard** is the default. You can also select **Verbose**. All verification messages (including Verbose messages) are logged to **/var/hacmp/clverify/clverify.log**.

3. Press Enter and SMIT starts the verification process. The verification output appears in the SMIT Command Status window.
4. If any error messages appear, make the necessary changes and run the verification procedure again. You may see Warnings if the configuration has a limitation on its availability; for example, only one interface per node per network is configured, or Workload Manager is configured but there is no application server assigned to use it.

Running Corrective Actions during Verification

You can run automatic corrective actions during cluster verification on an inactive cluster. By default, automatic corrective action is enabled for **Initialization and Standard** path and disabled for **Extended** path.

Automatic corrective actions can be disabled for the **Extended** path (from the **System Management (C-SPOC) > Manage HACMP Services >** menu) but it can*not* be disabled for the **Standard** path. You can run verification with corrective actions in one of two modes:

- *Interactively*. If you select **Interactively**, when verification detects a correctable condition related to importing a volume group or to re-importing mount points and filesystems, you are prompted to authorize a corrective action before verification continues.
- *Automatically*. If you select **Yes**, when verification detects that any of the error conditions exists, as listed in section Conditions That Can Trigger a Corrective Action, it takes the corrective action automatically without a prompt.

If an error discovered during verification has a corrective action, the item is corrected and the run continues. For situations when the correction involves importing a shared volume group, re-importing a shared volume group, or updating the **/etc/hosts** file, the utility runs all verification checks again after it corrects one of the above conditions. If the same error condition is triggered again, the associated corrective action is *not* executed. The error is logged and verification fails. If the original condition is a warning, verification succeeds.

HACMP 5.4 detects active service IP labels and active volume groups on nodes regardless of whether or *not* nodes are running cluster services. HACMP looks at the resource group state instead of cluster services. When verification detects active resources on a node that does *not* have the resource group in ONLINE state or does *not* have the resource group in an UNMANAGED state, verification gives you the option to bring these resources OFFLINE according the following table.

Verification can*not* tell which node will actually acquire the resource group that has active resources. Thus, the warning messages mentioned in the following table are printed every time active resources are found on a node that is or is *not* stopped and the state of the resource group to which active resources belong is UNMANAGED, OFFLINE, or ERROR.

	Resource Group Attribute: Manage Resource Group Automatically	**Resource Group Attribute: Manage Resource Group Manually**
Interactively correct errors	Display message with option to bring resources offline.	Display message with option to bring resources offline.
Automatically correct errors	Reset the startup attribute Managed Resource group to: Manually. Display warning message.	Print reminder/warning and steps to take.
No corrective actions	Print reminder/warning and steps to take	Print reminder/warning and steps to take
Cluster Services are running	N/A	N/A

Conditions That Can Trigger a Corrective Action

HACMP shared volume group time stamps are not up-to-date on a node

If the shared volume group time stamp file does *not* exist on a node, or the time stamp files do *not* match on all nodes, the corrective action ensures that all nodes have the latest up-to-date VGDA time stamp for the volume group and imports the volume group on all cluster nodes where the shared volume group was out of sync with the latest volume group changes. The corrective action ensures that volume groups whose definitions have changed will be properly imported on a node that does *not* have the latest definition.

The /etc/hosts file on a node does not contain all HACMP-managed IP addresses

If an IP label is missing, the corrective action modifies the file to add the entry and saves a copy of the old version to **/etc/hosts.*date***. If a backup file already exists for that day, no additional backups are made for that day.

- If the **/etc/hosts** entry exists but is commented out, verification adds a new entry; comment lines are ignored.
- If the label specified in the HACMP Configuration does *not* exist in **/etc/hosts**, but the IP address is defined in **/etc/hosts,** the label is added to the existing **/etc/hosts** entry. If the label is different between **/etc/hosts** and the HACMP configuration, then verification reports a different error message; no corrective action is taken.
- If the entry does *not* exist, meaning both the IP address and the label are missing from **/etc/hosts**, then the entry is added. This corrective action takes place on a node-by-node basis. If different nodes report different IP labels for the same IP address, verification catches these cases and reports an error. However, this error is unrelated to this corrective action. Inconsistent definitions of an IP label defined to HACMP are *not* corrected.

SSA concurrent volume groups need unique SSA node numbers

If verification finds that the SSA node numbers are *not* unique, the corrective action changes the number of one of the nodes where the number is *not* unique. See the *Installation Guide* for more information on SSA configuration.

Note: The SSA node number check is *not* performed for enhanced concurrent volume group that sit on the SSA hdisks. Disks that make up enhanced concurrent volume groups do *not* have any SSA-specific numbers assigned to them.

A filesystem is not created on a node, although disks are available

If a filesystem has *not* been created on one of the cluster nodes, but the volume group is available, the corrective action creates the mount point and filesystem. The filesystem must be part of a resource group for this action to take place. In addition, the following conditions must be met:

- This is a shared volume group.
- The volume group must already exist on at least one node.
- One or more node(s) that participate in the resource group where the filesystem is defined must already have the filesystem created.
- The filesystem must already exist within the logical volume on the volume group in such a way that simply re-importing that volume group would acquire the necessary filesystem information.

- The mount point directory must already exist on the node where the filesystem does *not* exist.

The corrective action handles only those mount points that are on a shared volume group, such that exporting and re-importing of the volume group will acquire the missing filesystems available on that volume group. The volume group is varied off on the remote node(s), or the cluster is down and the volume group is then varied off if it is currently varied on, prior to executing this corrective action.

If **Mount All Filesystems** is specified in the resource group, the node with the latest time stamp is used to compare the list of filesystems that exists on that node with other nodes in the cluster. If any node is missing a filesystem, then HACMP imports the filesystem.

Disks are available, but the volume group has not been imported to a node

If the disks are available but the volume group has *not* been imported to a node that participates in a resource group where the volume group is defined, then the corrective action imports the volume group.

The corrective action gets the information regarding the disks and the volume group major number from a node that already has the volume group available. If the major number is unavailable on a node, the next available number is used. The corrective action is only performed under the following conditions:

- The cluster is down.
- The volume group is varied off if it is currently varied on.
- The volume group is defined as a resource in a resource group.
- The major number and associated PVIDS for the disks can be acquired from a cluster node that participates in the resource group where the volume group is defined.

Note: This functionality will *not* turn off the **auto varyon** flag if the volume group has the attribute set. A separate corrective action handles auto varyon.

Shared volume groups configured as part of an HACMP resource group have their automatic varyon attribute set to Yes.

If verification finds that a shared volume group inadvertently has the auto varyon attribute set to **Yes** on any node, the corrective action automatically sets the attribute to **No** on that node.

Required /etc/services entries are missing on a node.

If a required entry is commented out, missing, or invalid in **/etc/services** on a node, the corrective action adds it. Required entries are:

Note: Starting with HACMP 5.3, the software no longer uses the **clsmuxpd** daemon; the SNMP server functions are included in the Cluster Manager—the **clstrmgr** daemon.

Name	Port	Protocol
topsvcs	6178	udp
grpsvcs	6179	udp

Name	Port	Protocol
clinfo_deadman	6176	udp
clcomd	6191	tcp

Required HACMP snmpd entries are missing on a node

If a required entry is commented out, missing, or invalid on a node, the corrective action adds it.

Note: The default version of the **snmpd.conf** file for AIX 5L v.5.2 and v. 5.3 is **snmpdv3.conf**.

In **/etc/snmpdv3.conf** or **/etc/snmpd.conf**, the required HACMP **snmpd** entry is:

```
smux   1.3.6.1.4.1.2.3.1.2.1.5 "clsmuxpd_password" # HACMP clsmuxpd
```

In **/etc snmpd.peers**, the required HACMP snmpd entry is:

```
clsmuxpd 1.3.6.1.4.1.2.3.1.2.1.5 "clsmuxpd_password" # HACMP clsmuxpd
```

If changes are required to the **/etc/snmpd.peers** or **snmpd[v3].conf** file, HACMP creates a backup of the original file. A copy of the pre-existing version is saved prior to making modifications in the file **/etc/snmpd.**{**peers | conf**}***.date***. If a backup has already been made of the original file, then no additional backups are made.

HACMP makes one backup per day for each **snmpd** configuration file. As a result, running verification a number of times in one day only produces one backup file for each file modified. If no configuration files are changed, HACMP does *not* make a backup.

Required RSCT Network Options Settings

HACMP requires that the **nonlocsrcroute**, **ipsrcroutesend**, **ipsrcrouterecv**, and **ipsrcrouteforward** network options be set to 1; these are set by RSCT's **topsvcs** startup script. The corrective action run on *inactive* cluster nodes ensures these options are *not* disabled and are set correctly.

Required HACMP Network Options Settings

The corrective action ensures that the value of each of the following network options is consistent across all nodes in a running cluster (out-of-sync setting on any node is corrected):

- **tcp_pmtu_discover**
- **udp_pmtu_discover**
- **ipignoreredirects**

Required routerevalidate Network Option Setting

Changing hardware and IP addresses within HACMP changes and deletes routes. Because AIX 5L caches routes, setting the **routerevalidate** network option is required as follows:

```
no -o routerevalidate=1
```

This setting ensures the maintenance of communication between cluster nodes. Verification run with corrective action automatically adjusts this setting for nodes in a running cluster.

Note: No corrective actions take place during a dynamic reconfiguration event.

clverify.log File

During verification, HACMP collects configuration data from all the nodes as it runs through a series of checks. The verbose output is saved to the **/var/hacmp/clverify/clverify.log** file. The log file is rotated; this helps you and IBM Support obtain a history of what configuration changes have been made when you need to determine the root cause of a problem.

Ten copies of the log are saved, as follow:

```
drwxr-xr-x   3 root     system         1024 Mar 13 00:02 .
drwxr-xr-x   6 root     system          512 Mar 11 10:03 ..
-rw-------   1 root     system       165229 Mar 13 00:02 clverify.log
-rw-------   1 root     system       165261 Mar 12 17:31 clverify.log.1
-rw-------   1 root     system       165515 Mar 12 15:22 clverify.log.2
-rw-------   1 root     system       163883 Mar 12 15:04 clverify.log.3
-rw-------   1 root     system       164781 Mar 12 14:54 clverify.log.4
-rw-------   1 root     system       164459 Mar 12 14:36 clverify.log.5
-rw-------   1 root     system       160194 Mar 12 09:27 clverify.log.6
-rw-------   1 root     system       160410 Mar 12 09:20 clverify.log.7
-rw-------   1 root     system       160427 Mar 12 09:16 clverify.log.8
-rw-------   1 root     system       160211 Mar 12 09:06 clverify.log.9
```

You can redirect the **clverify.log** file to write to a different location using the standard HACMP logfile redirection mechanism. If the **clverify.log** file is redirected to a different location, the location of all the data saved in the subdirectories in the path **/var/hacmp/clverify** moves along with it. However, pre-existing data under **/var/hacmp/clverify** is *not* automatically moved if the **clverify.log** is redirected.

For information on this procedure see the Steps for Redirecting a Cluster Log File section in Chapter 1: Using Cluster Log Files in the *Troubleshooting Guide*.

Archived Configuration Databases

All verification checks use HACMP Configuration Database data supplied by the common communication infrastructure, which is designed to provide efficient access to configuration databases from the other nodes. When the verification runs, it stores copies of the following:

- All HACMP Configuration Databases (ODMs) used during verification
- All AIX 5L ODMs (Custom Attributes, Device Definitions, and so forth) collected from the remote nodes.

The verification utility manages these files by storing the copies in various directories depending on the success or failure of the verification.

Managing HACMP File Collections

HACMP requires that event scripts, application scripts, AIX 5L files, and HACMP configuration files must be identical on each cluster node. The HACMP File Collections facility automatically synchronizes these files among cluster nodes and warns you if there are any unexpected results (for example, if one or more files in a collection has been deleted or has a length of zero on one or more cluster nodes).

Default HACMP File Collections

When you install HACMP, it sets up the following file collections:

- **Configuration_Files**
- **HACMP_Files**

HACMP Configuration_Files Collection

Configuration_Files is a container for the following essential system files:

- **/etc/hosts**
- **/etc/services**
- **/etc/snmpd.conf**
- **/etc/snmpdv3.conf**
- **/etc/rc.net**
- **/etc/inetd.conf**
- **/usr/es/sbin/cluster/netmon.cf**
- **/usr/es/sbin/cluster/etc/clhosts**
- **/usr/es/sbin/cluster/etc/rhosts**
- **/usr/es/sbin/cluster/etc/clinfo.rc**

For more information on the **/netmon.cf file** configuration see the *Planning Guide*, and for information about the **/clhosts** file during an upgrade, see the *Installation Guide.*

HACMP_Files Collection

HACMP_Files is a container for user-configurable files in the HACMP configuration. HACMP uses this file collection to reference all of the user-configurable files in the HACMP Configuration Database classes.

The **HACMP_Files** collection references the following Configuration Database fields:

Configuration Database Class	Configuration Database Field	Description
HACMPevent:	notify	Event notify script
HACMPevent:	pre	Pre-event script
HACMPevent:	post	Post-event script
HACMPevent:	recv	Recovery script
HACMPserver:	start	Application server start script
HACMPserver:	stop	Application server stop script
HACMPmonitor:	value, when name=NOTIFY_METHOD	Application monitor notify script

HACMPmonitor:	value, when name=CLEANUP_METHOD	Application monitor cleanup script
HACMPmonitor:	value, when name=RESTART_METHOD	Application monitor restart script
HACMPpager:	filename	Pager text message file
HACMPsna:	app_svc_file	SNA link start and stop scripts
HACMPx25:	app_svc_file	X.25 link start and stop scripts
HACMPtape:	start_script_name	Tape start script
HACMPtape:	stop_script_name	Tape stop script
HACMPude:	recovery_prog_path	User-defined event recovery program
HACMPcustom:	value	Custom snapshot method script

Note: This collection excludes the **HACMPevent:cmd** event script. Do *not* modify or rename the HACMP event script files. Also, do *not* include HACMP event scripts in any HACMP file collection.

Note: When copying a file to a remote node, the local node's owner, group, modification time stamp, and permission settings are maintained on the remote node. That is, the remote node inherits these settings from the local node.

Permissions for all files in the **HACMP_Files** collection are set to *execute*, which helps to prevent problems if you have *not* yet set execute permission for scripts on all nodes. (This is often the cause of an event failure.)

You can*not* rename or delete the **HACMP_Files** collection. You can*not* add or remove files from the collection.

You can add a file that is already included in the **HACMP_Files** collection (for example, an application start script) to another file collection. However, in any other case, a file can only be included in *one* file collection and you receive the following error message, where XXX _Files is the name of the previously defined collection:

```
This file is already included in the <XXX_Files> collection).
```

You can add and remove files or delete the **Configuration_Files** collection.

Neither of these file collections is enabled by default. If you prefer to include some user-configurable files in another collection instead of propagating all of them, leave the **HACMP_Files** collection disabled.

Options for Propagating an HACMP File Collection

Propagating a file collection copies the files in a file collection from the current node to the other cluster nodes. Use one of the following methods to propagate an HACMP file collection:

- Propagate the file collection at any time manually. You can propagate files in a file collection from the HACMP File Collection SMIT menu on the local node (the node that has the files you want to propagate).
- Set the option to propagate the file collection whenever cluster verification and synchronization is executed. The node from which verification is run is the propagation node. (This is set to **No** by default.)
- Set the option to propagate the file collection automatically after a change to one of the files in the collection. HACMP checks the file collection status on each node (every 10 minutes by default) and propagates any changes. (This is set to **No** by default.)

  One timer is set for all file collections. You can change the timer. The maximum is 1440 minutes (24 hours) and the minimum is 10 minutes.

You can set up and change file collections on a running cluster. However, note that if you add a node dynamically, the file collection on that node may have files that are *not* in sync with the files on the other cluster nodes. If the file collection on the node being added is set for automatic propagation upon cluster verification and synchronization, the files on the node just added are updated properly. If this flag is *not* set, you must manually run the file collection propagation from one of the other nodes.

Backup Files and Error Handling

During file propagation, before HACMP copies a file to a remote node, the remote node makes a backup copy of the original file if it exists and its size is greater than zero, with the original time stamp. The copy is kept in the **/var/hacmp/filebackup/** directory.

Only the most recent backup is kept for each file that is overwritten. When another propagation replaces the file, the new backup overwrites the old one. You can*not* customize these backups. If you need to use a backup file, you must manually copy the file back to its original location.

If the local (propagation) node has a zero-length or non-existent file in a file collection, then an error message is logged and the file is *not* copied during the propagation process. The zero-length or non-existent file remains until you run a manual propagation from another node, or when an automatic propagation from another node sees a change to the file and propagates it.

All errors during file propagation are logged to SMIT if the propagation happens during a cluster verification or synchronization or manual propagation. Errors are also written to the **/var/hacmp/log/clutils.log** file.

It is your responsibility to ensure that the file on the local (propagation) node is the latest copy and is *not* corrupt. HACMP only checks for the existence and length of the file on this node.

Tracking HACMP File Collection Operations

Whenever the HACMP File Collections utility replaces a file on a node, the following information about it is saved in the **/var/hacmp/log/clutils.log** file:

- Date and time of replacement
- Propagation type

- File name and file collection name
- Name of the remote and local nodes.

For example:

```
Wed Jan 07 11:08:55 2006: clfileprop: Manual file collection propagation
called.
Wed Jan 07 11:08:55 2006: clfileprop: The following file collections
will be processed:
Wed Jan 07 11:08:55 2006: clfileprop: Test_Files Wed Jan 07 11:08:55
2004: clfileprop:
Wed Jan 07 11:08:55 2006: clfileprop: Starting file propagation to
remote node riga.
Wed Jan 07 11:08:55 2006: clfileprop: Successfully propagated file
/tmp/kris to node riga.
Wed Jan 07 11:08:55 2006: clfileprop: Successfully propagated file
/tmp/k2 to node riga.
Wed Jan 07 11:08:55 2006: clfileprop: Total number of files propagated
to node riga: 2
```

Using SMIT to Manage HACMP File Collections

The SMIT interface enables you to perform the following actions:

- Creating an HACMP File Collection
- Setting the Automatic Timer for File Collections
- Changing a File Collection
- Removing Files from a File Collection
- Removing a File Collection
- Verifying and Synchronizing File Collections.

Creating an HACMP File Collection

To create an HACMP File Collection, at least one working IP communications path defined to HACMP must exist between the node running the file propagation and each remote node defined to the cluster. The **clcomd** daemon must be running on all nodes.

To create an HACMP file collection:

1. Enter `smit hacmp`
2. In SMIT, select **System Management (C-SPOC) > HACMP File Collection Management > Manage File Collections > Add a File Collection** and press Enter.
3. Enter field values as follows:

File Collection name	The name can include alphabetic and numeric characters and underscores. Use no more than 32 characters. Do *not* use reserved names. For a list of reserved names, see List of Reserved Words.
File Collection Description	A description of the file collection. Use no more than 100 characters.

Propagate files during cluster synchronization?	**No** is the default. If you select **Yes**, HACMP propagates files listed in the current collection before every cluster verification and synchronization process.
Propagate changes to files automatically?	**No** is the default. If you select **Yes**, HACMP propagates files listed in the current collection across the cluster when a change is detected on any file in the collection. HACMP checks for changes every ten minutes by default. You can adjust the timer on the **Manage File Collections** panel.

4. In SMIT, select **HACMP File Collection Management > Manage Files in File Collections > Add Files to a File Collection** and press Enter.
5. Select the File Collection where you want to add the files.
6. Enter the file names in the **New Files** field:

File Collection name	The name of the selected file collection is displayed.
File Collection Description	The current description is displayed.
Propagate files during cluster synchronization?	The current choice is displayed.
Propagate changes to files automatically?	The current choice is displayed.
Collection Files	Any files already in the collection are displayed.
New Files	Add the full pathname of the new file. The name must begin with a forward slash. A file can*not* be a symbolic link, a directory, a pipe, a socket, or any file in **/dev** or **/proc**. It can*not* begin with **/etc/objrepos/*** or **/etc/es/objrepos/***. The file can*not* be in another file collection (except for **HACMP_Files**).

7. When you finish creating the file collection(s), synchronize the cluster using SMIT **Extended Configuration > Extended Verification and Synchronization**.

Setting the Automatic Timer for File Collections

The default timer for automatic checks on file collections is ten minutes. You can change the amount of time as needed.

Note: The periodic check for changes to a file in a file collection runs on each node. However, these checks are *not* coordinated to run simultaneously on every node. Make changes to a file *only on one node* within the general time limit.

To customize the file collection time interval:

1. Enter `smit hacmp`

2. In SMIT, select **System Management (C-SPOC) > HACMP File Collection Management > Manage File Collections > Change/Show Automatic Update Time** and press Enter.
3. Enter the amount of time (in minutes) that you want HACMP to pause before performing file collection synchronization. The maximum is 1440 minutes (24 hours) and the minimum is 10 minutes. Press Enter.
4. Synchronize the cluster using SMIT (**Extended Configuration > Extended Verification and Synchronization**).

Changing a File Collection

You can modify a file collection as follows:

- Change the attributes of a file collection (name, description, propagation parameters).
- Add or remove files in the collection.
- Remove a file collection.
- Change the automatic timer for all file collections, as described in Setting the Automatic Timer for File Collections.

To change an attribute of a particular file collection:

1. Enter `smit hacmp`
2. In SMIT, select **System Management (C-SPOC) > HACMP File Collection Management > Manage File Collections > Change/Show a File Collection** and press Enter.
3. Select the file collection.
4. Change the name, description, and synchronization parameters on this panel:

File Collection name	The current name appears here.
New File Collection name	Enter the new name.
Propagate files during cluster synchronization?	**No** is the default. If you select **Yes**, HACMP propagates files listed in the current collection before every cluster verification and synchronization process.
Propagate changes to files automatically?	**No** is the default. If you select **Yes**, HACMP propagates files listed in the current collection across the cluster automatically when a change is detected on any file in the collection. HACMP checks for changes every ten minutes by default. You can adjust the timer on the **Manage File Collections** panel.
Collection Files	Any files already in the collection are displayed. Press F4 to see the list. You can*not* change this field.

5. Synchronize the cluster. In SMIT, select **Extended Configuration > Extended Verification and Synchronization** and press Enter.

Removing Files from a File Collection

To remove files from a file collection:

1. Enter `smit hacmp`
2. In SMIT, select **System Management (C-SPOC) > HACMP File Collection Management > Manage Files in File Collections** > **Remove Files from a File Collection** and press Enter.
3. Select the File Collection from which you want to remove the files.
4. Select one or more files to remove from the file collection and press Enter.
5. Synchronize the cluster to update Configuration Databases. In SMIT, select **Extended Configuration > Extended Verification and Synchronization** and press Enter.

Removing a File Collection

To remove a file collection from the HACMP configuration:

1. Enter `smit hacmp`
2. In SMIT, select **System Management (C-SPOC) > HACMP File Collection Management > Manage File Collections** > **Remove a File Collection** and press Enter.
3. Select the file collection to remove and press Enter.
4. SMIT displays `Are you sure?` Press Enter again.
5. In SMIT, select **Extended Configuration > Extended Verification and Synchronization** and press Enter to synchronize the cluster.

Verifying and Synchronizing File Collections

If file collections exist, HACMP checks and propagates the file collections with the flag set to **yes** for “propagate during verify and synchronize” before running the rest of the cluster verification and synchronization process. Before the files in each collection are propagated to all the cluster nodes, HACMP performs the following verification checks:

- Verifies that no files are listed twice in any file collection. If a file is listed twice, a warning displays and verification continues.
- Verifies that each file listed in each collection is a real file on the local node (the node from which cluster synchronization is being run). A file can*not* be a symbolic link, a directory, a pipe, a socket, or any file in **/dev** or **/proc**. It can*not* begin with **/etc/objrepos/*** or **/etc/es/objrepos/***. If a file in a file collection is one of these, HACMP displays an error and verification fails.
- Verifies that each file exists on the local node and has a file size greater than zero. If a file does *not* exist on the local node or has a size of zero, HACMP displays an error and verification fails.
- Verifies that each file has a full path name that begins with a forward slash. If a file's pathname does *not* begin with a forward slash, HACMP displays an error and verification fails.

Adding a Custom Verification Method

You may want to add a custom verification method to check for a particular issue on your cluster. For example, you could add a script to check for the version of an application. You could include an error message for display and to write to the **clverify.log** file.

Note: During node startup, automatic verification and synchronization does *not* include any custom verification methods.

To add a custom verification method:

1. Enter `smit hacmp`
2. In SMIT, select **Problem Determination Tools** > **HACMP Verification** > **Configure Custom Verification Method** > **Add a Custom Verification Method** and press Enter**.**
3. Enter the field values as follows:

Verification Method Name	Enter a name for the verification method. Method names can be up to 32 alphanumeric characters. Do *not* use the word "all," as this is a keyword indicating that all custom verification methods are to be run.
Verification Method Description	Enter a short description of the verification method.
Verification Method	Enter a filename for the verification method (executable). The method name can be different from the filename.

4. Press Enter. The method is added to the list of verification methods you can use when you select the HACMP Verification option under the **Problem Determination Tools** menu.

Changing or Showing a Custom Verification Method

To change or show a custom verification method:

1. Enter `smit hacmp`
2. From the **Problem Determination Tools** menu, select **HACMP Verification > Define Custom Verification Method > Change/Show a Custom Verification Method** and press Enter. SMIT displays a popup list of verification methods
3. Select the verification method you want to change or show and press Enter.
4. Enter a new name, new verification method description, and/or new filename as desired for the verification method and press Enter.

Removing a Custom Verification Method

To remove a custom verification method:

1. Enter `smit hacmp`

2. In SMIT, select **Problem Determination Tools** menu, select **HACMP Verification > Define Custom Verification Method > Remove a Custom Verification Method** and press Enter. SMIT displays a popup list of custom verification methods.
3. Select the verification method you want to remove and press Enter. SMIT prompts you to confirm that you want to remove the specified verification method.
4. Press Enter to remove the verification method.

List of Reserved Words

Do *not* use the following words as names in a cluster. However, you may use these words when combined with numerals or another word (for example, `my_network` or `rs232_02`).

adapter	false	nim	socc
alias	FBHPN	node	subnet
all	fcs	nodename	tmscsi
ALL	fddi	OAAN	tmssa
ANY	FNPN	OFAN	token
atm	fscsi	OHN	true
BO	FUDNP	OTHER	tty
cluster	grep	OUDP	volume
command	group	private	vpath
CROSS_SITE_RG_MOVE	hps	public	vscsi
custom	ib	resource	XD_data
daemon	ip	RESTORE	XD_ip
disk	IP	root	XD_rs232
diskhb	name	rs232	
ether	network	serial	
event	NFB	slip	

Chapter 8: Testing an HACMP Cluster

This chapter describes how to use the Cluster Test Tool to test the recovery capabilities of an HACMP cluster. The Cluster Test Tool is available for you to test a new cluster before it becomes part of your production environment, and to test configuration changes to an existing cluster, when the cluster is *not* in service.

The main sections of the chapter include:

Prerequisites

The Cluster Test Tool runs only on a cluster that has:

- HACMP 5.2 or greater installed
- If the cluster is migrated from an earlier version, cluster migration must be complete.
- If you used the Cluster Test Tool in previous releases, the custom test plans that you created in previous releases continue to work in HACMP v.5.4.
- The cluster configuration verified and synchronized.

Before you run the tool on a cluster node, ensure that:

- The node has HACMP installed and is part of the HACMP cluster to be tested.
- The node has network connectivity to all of the other nodes in the HACMP cluster.
- You have root permissions.

Because log file entries include time stamps, consider synchronizing the clocks on the cluster nodes to make it easier to review log file entries produced by test processing.

Overview

The Cluster Test Tool utility lets you test an HACMP cluster configuration to evaluate how a cluster operates under a set of specified circumstances, such as when cluster services on a node fail or when a node loses connectivity to a cluster network. You can start a test, let it run unattended, and return later to evaluate the results of your testing. You should run the tool under both low load and high load conditions to observe how system load affects your HACMP cluster.

You run the Cluster Test Tool from SMIT on one node in an HACMP cluster. For testing purposes, this node is referred to as the *control node*. From the control node, the tool runs a series of specified tests—some on other cluster nodes, gathers information about the success or failure of the tests processed, and stores this information in the Cluster Test Tool log file for evaluation or future reference.

The Cluster Test Tool lets you test an HACMP cluster in two ways, by running:

- Automated testing (also known as Automated Test Tool). In this mode, the Cluster Test Tool runs a series of predefined sets of tests on the cluster.
- Custom testing (also known as Test Plan). In this mode, you can create your own test plan, or a custom testing routine, that will include different tests available in the Cluster Test Tool library.

Automated Testing

Use the automated test procedure (a predefined set of tests) supplied with the tool to perform basic cluster testing on any cluster. No setup is required. You simply run the test from SMIT and view test results from SMIT and the Cluster Test Tool log file.

The automated test procedure runs a predefined set of tests on a node that the tool randomly selects. The tool ensures that the node selected for testing varies from one test to another. For information about automated testing, see the section Running Automated Tests.

Custom Testing

If you are an experienced HACMP administrator and want to tailor cluster testing to your environment, you can create custom tests that can be run from SMIT. You create a custom test plan (a file that lists a series of tests to be run), to meet requirements specific to your environment and apply that test plan to any number of clusters. You specify the order in which tests run and the specific components to be tested. After you set up your custom test environment, you run the test procedure from SMIT and view test results in SMIT and in the Cluster Test Tool log file. For information about customized testing, see the section Setting up Custom Cluster Testing.

Test Duration

Running automated testing on a basic two-node cluster that has a simple cluster configuration takes approximately 30 to 60 minutes to complete. Individual tests can take around three minutes to run. The following conditions affect the length of time to run the tests:

- Cluster complexity

  Testing in complex environments takes considerably longer.

- Latency on the network

  Cluster testing relies on network communication between the nodes. Any degradation in network performance slows the performance of the Cluster Test Tool.
- Use of verbose logging for the tool

  If you customize verbose logging to run additional commands from which to capture output, testing takes longer to complete. In general, the more commands you add for verbose logging, the longer a test procedure takes to complete.
- Manual intervention on the control node

  At some points in the test, you may need to intervene. See Recovering the Control Node after Cluster Manager Stops for ways to avoid this situation.
- Running custom tests

  If you run a custom test plan, the number of tests run also affects the time required to run the test procedure. If you run a long list of tests, or if any of the tests require a substantial amount of time to complete, then the time to process the test plan increases.

Security

The Cluster Test Tool uses the HACMP Cluster Communications daemon to communicate between cluster nodes to protect the security of your HACMP cluster. For information about the Cluster Communications Daemon, see Chapter 16: Managing User and Groups.

Limitations

The Cluster Test Tool has the following limitations. It does *not* support testing of the following HACMP cluster-related components:

- High Performance Switch (HPS) networks
- ATM networks
- Sites.

  You can perform general cluster testing for clusters that support sites, but *not* testing specific to HACMP sites or any of the HACMP/XD products. HACMP/XD for Metro Mirror HACMP/XD for GLVM, and HACMP/XD for HAGEO all use sites in their cluster configuration.
- Replicated resources.

  You can perform general cluster testing for clusters that include replicated resources, but *not* testing specific to replicated resources or any of the HACMP/XD products. HACMP/XD for Metro Mirror, HACMP/XD for HAGEO, and HACMP/XD for GLVM all include replicated resources in their cluster configuration.
- Dynamic cluster reconfiguration.

  You can*not* run dynamic reconfiguration while the tool is running.
- Pre-events and post-events.

  Pre-events and post-events run in the usual way, but the tool does *not* verify that the events were run or that the correct action was taken.

In addition, the Cluster Test Tool may *not* recover from the following situations:

- A node that fails unexpectedly, that is a failure *not* initiated by testing

- The cluster does *not* stabilize.

Note: The Cluster Test Tool uses the terminology for stopping cluster services that was used in HACMP prior to v.5.4 (graceful stop, graceful with takeover and forced stop). For information how this terminology maps to the currently used terms for stopping the cluster services, see Chapter 9: Starting and Stopping Cluster Services.

Running Automated Tests

You can run the automated test procedure on any HACMP cluster that is *not* currently in service. The Cluster Test Tool runs a specified set of tests and randomly selects the nodes, networks, resource groups, and so forth for testing. The tool tests different cluster components during the course of the testing. For a list of the tests that are run, see the section Understanding Automated Testing.

Before you start running an automated test:

- Ensure that the cluster is *not* in service in a production environment
- Stop HACMP cluster services, this is recommended but optional. Note that if the Cluster Manager is running, some of the tests will be irrational for your configuration, but the Test Tool will continue to run.
- Cluster nodes are attached to two IP networks.

  One network is used to test a network becoming unavailable then available. The second network provides network connectivity for the Cluster Test Tool. Both networks are tested, one at a time.

Launching the Cluster Test Tool

To run the automated test procedure:

1. Enter `smit hacmp`
2. In SMIT, select **Initialization and Standard Configuration > HACMP Cluster Test Tool** and press Enter.

   The `Are you sure` message appears. If you press Enter again, the automated test plan runs.
3. Evaluate the test results.

   For information about evaluating test results, see the section Evaluating Results.

Modifying Logging and Stopping Processing in the Cluster Test Tool

You can also modify processing for automated test procedure to:

- Turn off verbose logging
- Turn off cycling of log files for the tool
- Stop processing tests after the first test fails

To modify processing for an automated test:

1. Enter `smit hacmp`
2. In SMIT, select either one of the following options:
   - **Extended Configuration**
   - **Problem Determination Tools**

   Then select **HACMP Cluster Test Tool.**
3. In the **HACMP Cluster Test Tool** panel, select **Execute Automated Test Procedure**.
4. In the **Execute Automated Test Procedure** panel, enter field values as follows:

Verbose Logging	When set to **yes**, includes additional information in the log file. This information may help to judge the success or failure of some tests. For more information about verbose logging and how to modify it for your testing, see the section Error Logging. Select **no** to decrease the amount of information logged by the Cluster Test Tool. The default is **yes**.
Cycle Log File	When set to **yes,** uses a new log file to store output from the Cluster Test Tool. Select **no** to append messages to the current log file. The default is **yes**. For more information about cycling the log file, see the section Log File Rotation.
Abort on Error	When set to **no,** the Cluster Test Tool continues to run tests after some of the tests being run fail. This may cause subsequent tests to fail because the cluster state is different from the one expected by one of those tests. Select **yes** to stop processing after the first test fails. For information about the conditions under which the Cluster Test Tool stops running, see the section Cluster Test Tool Stops Running. The default is **no**. **Note:** The tool stops running and issues an error if a test fails and **Abort on Error** is selected.

5. Press Enter to start running the automated tests.
6. Evaluate the test results.

   For information about evaluating test results, see the section Evaluating Results.

Understanding Automated Testing

This section lists the sequence that the Cluster Test Tool uses for the automated testing, and describes the syntax of the tests run during automated testing.

The automated test procedure performs sets of predefined tests in the following order:

1. General topology tests
2. Resource group tests on non-concurrent resource groups
3. Resource group tests on concurrent resource groups
4. IP-type network tests for each network
5. Non-IP network tests for each network
6. Volume group tests for each resource group
7. Site-specific tests
8. Catastrophic failure test.

The Cluster Test Tool discovers information about the cluster configuration, and randomly selects cluster components, such as nodes and networks, to be used in the testing.

Which nodes are used in testing varies from one test to another.The Cluster Test Tool may select some node(s) for the initial battery of tests, and then, for subsequent tests, it may intentionally select the same node(s), or, choose from nodes on which no tests were run previously. In general, the logic in the automated test sequence ensures that all components are sufficiently tested in all necessary combinations. The testing follows these rules:

- Tests operation of a concurrent resource group on one randomly selected node—*not* all nodes in the resource group.
- Tests only those resource groups that include monitored application servers or volume groups.
- Requires at least two active IP networks in the cluster to test non-concurrent resource groups.

The automated test procedure runs a **node_up** event at the beginning of the test to make sure that all cluster nodes are up and available for testing.

These sections list the tests in each group. For more information about a test, including the criteria to determine the success or failure of a test, see the section Description of Tests. The automated test procedure uses variables for parameters, with values drawn from the HACMP cluster configuration.

The examples in the following sections use variables for node, resource group, application server, stop script, and network names. For information about the parameters specified for a test, see the section Description of Tests.

General Topology Tests

The Cluster Test Tool runs the general topology tests in the following order:

1. Bring a node up and start cluster services on all available nodes
2. Stop cluster services on a node and bring resource groups offline.
3. Restart cluster services on the node that was stopped
4. Stop cluster services and move resource groups to another node
5. Restart cluster services on the node that was stopped

6. Stop cluster services on another node and place resource groups in an UNMANAGED state.
7. Restart cluster services on the node that was stopped.

The Cluster Test Tool uses the terminology for stopping cluster services that was used in HACMP in releases prior to v.5.4. For information on how the methods for stopping cluster services map to the terminology used in v.5.4, see Chapter 9: Starting and Stopping Cluster Services.

When the automated test procedure starts, the tool runs each of the following tests in the order shown:

1. `NODE_UP, ALL, Start cluster services on all available nodes`
2. `NODE_DOWN_GRACEFUL, node1, Stop cluster services gracefully on a node`
3. `NODE_UP, node1, Restart cluster services on the node that was stopped`
4. `NODE_DOWN_TAKEOVER, node2, Stop cluster services with takeover on a node`
5. `NODE_UP, node2, Restart cluster services on the node that was stopped`
6. `NODE_DOWN_FORCED, node3, Stop cluster services forced on a node`
7. `NODE_UP, node3, Restart cluster services on the node that was stopped`

Resource Group Tests

There are two groups of resource group tests that can be run. Which group of tests run depends on the startup policy for the resource group: non-concurrent and concurrent resource groups.

If a resource of the specified type does *not* exist in the resource group, the tool logs an error in the Cluster Test Tool log file.

Resource Group Starts on a Specified Node

The following tests run if the cluster includes one or more resource groups that have a startup management policy *other than* Online on All Available Nodes, that is, the cluster includes one or more non-concurrent resource groups.

The Cluster Test Tool runs each of the following tests in the order shown for each resource group:

1. Bring a resource group offline and online on a node.

   `RG_OFFLINE, RG_ONLINE`
2. Bring a local network down on a node to produce a resource group fallover.

   `NETWORK_DOWN_LOCAL, rg_owner, svc1_net, Selective fallover on local network down`
3. Recover the previously failed network.

```
NETWORK_UP_LOCAL, prev_rg_owner, svc1_net, Recover previously failed
network
```

4. Move a resource group to another node. `RG_MOVE`
5. Bring an application server down and recover from the application failure.
   `SERVER_DOWN, ANY, app1, /app/stop/script, Recover from application failure`

Resource Group Starts on All Available Nodes

If the cluster includes one or more resource groups that have a startup management policy of **Online on All Available Nodes**, that is, the cluster has concurrent resource groups, the tool runs one test that brings an application server down and recovers from the application failure.

The tool runs the following test:

```
RG_OFFLINE, RG_ONLINE
SERVER_DOWN, ANY, app1, /app/stop/script, Recover from
application failure
```

Network Tests

The tool runs tests for IP networks and for non-IP networks.

For each IP network, the tool runs these tests:

- Bring a network down and up.

  `NETWORK_DOWN_GLOBAL, NETWORK_UP_GLOBAL`
- Fail a network interface, join a network interface. This test is run for the service interface on the network. If no service interface is configured, the test uses a random interface defined on the network.

  `FAIL_LABEL, JOIN_LABEL`

For each Non-IP network, the tool runs these tests:

- Bring a non-IP network down and up.

  `NETWORK_DOWN_GLOBAL, NETWORK_UP_GLOBAL`

Volume Group Tests

For each resource group in the cluster, the tool runs tests that fail a volume group in the resource group: `VG_DOWN`

Site-Specific Tests

If sites are present in the cluster, the tool runs tests for them. The automated testing sequence that the Cluster Test Tool uses contains two site-specific tests:

- `auto_site`. This sequence of tests runs if you have *any* cluster configuration with sites. For instance, this sequence is used for clusters with cross-site LVM mirroring configured that does not use XD_data networks. The tests in this sequence include:
  - SITE_DOWN_GRACEFUL Stop the cluster services on all nodes in a site while taking resources offline
  - SITE_UP Restart the cluster services on the nodes in a site

- SITE_DOWN_TAKEOVER Stop the cluster services on all nodes in a site and move the resources to nodes at another site
- SITE_UP Restart the cluster services on the nodes at a site
- RG_MOVE_SITE Move a resource group to a node at another site

- `auto_site_isolation`. This sequence of tests runs only if you configured sites *and* an XD-type network. The tests in this sequence include:
  - SITE_ISOLATION Isolate sites by failing XD_data networks
  - SITE_MERGE Merge sites by bringing up XD_data networks.

Catastrophic Failure Test

As a final test, the tool stops the Cluster Manager on a randomly selected node that currently has at least one active resource group:

```
CLSTRMGR_KILL, node1, Kill the cluster manager on a node
```

If the tool terminates the Cluster Manager on the control node, you may need to reboot this node.

Setting up Custom Cluster Testing

If you want to extend cluster testing beyond the scope of the automated testing and you are an experienced HACMP administrator who has experience planning, implementing, and troubleshooting clusters, you can create a custom test procedure to test the HACMP clusters in your environment. You can specify the tests specific to your clusters, and use variables to specify parameters specific to each cluster. Using variables lets you extend a single custom test procedure to run on a number of different clusters. You the run the custom test procedure from SMIT.

WARNING: If you uninstall HACMP, the program removes any files you may have customized for the Cluster Test Tool. If you want to retain these files, make a copy of these files before you uninstall HACMP.

Planning a Test Procedure

Before you create a test procedure, make sure that you are familiar with the HACMP clusters on which you plan to run the test. List the components in your cluster and have this list available when setting up a test. Include the following items in the list:

- Nodes
- IP networks
- Non-IP networks
- XD-type networks
- Volume groups
- Resource groups
- Application servers

- Sites.

Your test procedure should bring each component offline then online, or cause a resource group fallover, to ensure that the cluster recovers from each failure.

We recommend that your test start by running a **node_up** event on each cluster node to ensure that all cluster nodes are up and available for testing.

Creating a Custom Test Procedure

To create a custom test procedure:

1. Create a Test Plan, a file that lists the tests to be run.

   For information about creating a Test Plan, see the section Creating a Test Plan.

2. Set values for test parameters.

   For information about specifying parameters, see the section Specifying Parameters for Tests.

Creating a Test Plan

A Test Plan is a text file that lists cluster tests to be run in the order in which they are listed in the file. In a Test Plan, specify one test per line. You can set values for test parameters in the Test Plan or use variables to set parameter values.

The tool supports the following tests:

FAIL_LABEL	Brings the interface associated with the specified label down on the specified node.
JOIN_LABEL	Brings the interface associated with the specified label up on the specified node.
NETWORK_UP_GLOBAL	Brings a specified network up (IP network or non-IP network) on all nodes that have interfaces on the network.
NETWORK_DOWN_GLOBAL	Brings a specified network down (IP network or non-IP network) on all nodes that have interfaces on the network.
NETWORK_UP_LOCAL	Brings a network interface on a node up.
NETWORK_DOWN_LOCAL	Brings a network interface on a node down.
NETWORK_UP_NONIP	Brings a non-IP network on a node up.
NETWORK_DOWN_NONIP	Brings a non-IP network on a node down.
NODE_UP	Starts cluster services on the specified node.
NODE_DOWN_GRACEFUL	Stops cluster services and brings the resource groups offline on the specified node.

NODE_DOWN_TAKEOVER	Stops cluster services with the resources acquired by another node.
NODE_DOWN_FORCED	Stops cluster services on the specified node with the Unmanage Resource Group option.
CLSTRMGR_KILL	Terminates the Cluster Manager on the specified node
RG_MOVE	Moves a resource group that is already online to a specific node
RG_MOVE_SITE	Moves a resource group that is already online to an available node at a specific site.
RG_OFFLINE	Brings a resource group offline that is already online
RG_ONLINE	Brings a resource group online that is already offline
SERVER_DOWN	Brings a monitored application server down
SITE_ISOLATION	Brings down all XD_data networks in the cluster at which the tool is running, thereby causing a site isolation.
SITE _MERGE	Brings up all XD_data networks in the cluster at which the tool is running, thereby simulating a site merge. Run the SITE_MERGE test after running the SITE_ISOLATION test.
SITE_UP	Starts cluster services on all nodes at the specified site that are currently stopped
SITE_DOWN_TAKEOVER	Stops cluster services on all nodes at the specified site and moves the resources to node(s) at another site by launching automatic **rg_move** events.
SITE_DOWN_GRACEFUL	Stops cluster services on all nodes at the specified site and takes the resources offline.
VG_DOWN	Emulates an error condition for a specified disk that contains a volume group in a resource group.
WAIT	Generates a wait period for the Cluster Test Tool.

For a full description of these tests, see the section Description of Tests.

Specifying Parameters for Tests

You can specify parameters for the tests in the Test Plan by doing one of the following:

- Using a variables file. A variables file defines values for variables assigned to parameters in a test plan. See the section Using a Variables File.
- Setting values for test parameters as environment variables. See the section Using Environment Variables.
- Identifying values for parameters in the Test Plan. See the section Using the Test Plan.

When the Cluster Test Tool starts, it uses a variables file if you specified the location of one in SMIT. If it does *not* locate a variables file, it uses values set in an environment variable. If a value is *not* specified in an environment variable, it uses the value in the Test Plan. If the value set in the Test Plan is *not* valid, the tool displays an error message.

Using a Variables File

The variables file is a text file that defines the values for test parameters. By setting parameter values in a separate variables file, you can use your Test Plan to test more than one cluster.

The entries in the file have this syntax:

parameter_name=value

For example, to specify a node as **node_waltham**:

```
node=node_waltham
```

To provide more flexibility, you can:

1. Set the name for a parameter in the Test Plan.
2. Assign the name to another value in the variables file.

For example, you could specify the value for *node* as **node1** in the Test Plan:

```
NODE_UP,node1, Bring up node1
```

In the variables file, you can then set the value of **node1** to **node_waltham**:

```
node1=node_waltham
```

The following example shows a sample variables file:

```
node1=node_waltham
node2=node_belmont
node3=node_watertown
node4=node_lexington
```

Using Environment Variables

If you do *not* want to use a variables file, you can assign parameter values by setting environment variables for the parameter values. If a variable file is *not* specified, but there are *parameter_name*=values in the cluster environment that match the values in the test plan, the Cluster Test Tool will use the values from the cluster environment.

Using the Test Plan

If you want to run a test plan on only one cluster, you can define test parameters in the Test Plan. The associated test can be run only on the cluster that includes those cluster attributes specified. For information about the syntax for parameters for tests, see the section Description of Tests.

Description of Tests

The Test Plan supports the tests listed in this section. The description of each test includes information about the test parameters and the success indicators for a test.

Note: One of the success indicators for each test is that the cluster becomes stable. The definition of cluster stability takes a number of factors into account, beyond the state of the Cluster Manager. The **clstat** utility, by comparison, uses only the state of the Cluster Manager to assess stability. For information about the factors used to determine cluster stability for the Cluster Test Tool, see the section Evaluating Results.

Test Syntax

The syntax for a test is:

TEST_NAME, *parameter1*, *parametern*|*PARAMETER*, *comments*

where:

- The test name is in uppercase letters.
- Parameters follow the test name.
- Italic text indicates parameters expressed as variables.
- Commas separate the test name from the parameters and the parameters from each other. (Note that the HACMP 5.4 Cluster Test Tool supports spaces around commas).

  The example syntax line shows parameters as *parameter1* and *parametern* with *n* representing the next parameter. Tests typically have from two to four parameters.
- A pipe (|) indicates parameters that are mutually exclusive alternatives.

  Select one of these parameter options.
- (*Optional*) Comments (user-defined text) appear at the end of the line. The Cluster Test Tool displays the text string when the Cluster Test Tool runs.

In the test plan, the tool ignores:

- Lines that start with a pound sign (#)
- Blank lines.

Node Tests

The node tests start and stop cluster services on specified nodes.

NODE_UP, *node* | ALL, *comments*

Starts cluster services on a specified node that is offline or on all nodes that are offline.

node	The name of a node on which cluster services start
ALL	Any nodes that are offline have cluster services start
comments	User-defined text to describe the configured test.

Example
```
NODE_UP, node1, Bring up node1
```

Entrance Criteria
Any node to be started is inactive.

Success Indicators
The following conditions indicate success for this test:

- The cluster becomes stable
- The cluster services successfully start on all specified nodes
- No resource group enters the error state
- No resource group moves from online to offline.

NODE_DOWN_GRACEFUL, *node* | ALL, *comments*

Stops cluster services on a specified node and brings resource groups offline.

node	The name of a node on which cluster services stop
ALL	All nodes are to have cluster services stop If you specify **ALL**, at least one node in the cluster must be online for this test to run.
comments	User-defined text to describe the configured test.

Example
```
NODE_DOWN_GRACEFUL, node3, Bring down node3 gracefully
```

Entrance Criteria
Any node to be stopped is active.

Success Indicators
The following conditions indicate success for this test:

- The cluster becomes stable
- Cluster services stop on the specified node(s)
- Cluster services continue to run on other nodes if **ALL** is *not* specified
- Resource groups on the specified node go offline, and do *not* move to other nodes
- Resource groups on other nodes remain in the same state.

NODE_DOWN_TAKEOVER, *node*, *comments*

Stops cluster services on a specified node with a resource group acquired by another node as configured, depending on resource availability.

node	The name of a node on which to stop cluster services
comments	User-defined text to describe the configured test.

Example
```
NODE_DOWN_TAKEOVER, node4, Bring down node4 gracefully with takeover
```

Entrance Criteria
The specified node is active.

Success Indicators
The following conditions indicate success for this test:

- The cluster becomes stable
- Cluster services stop on the specified node
- Cluster services continue to run on other nodes
- All resource groups remain in the same state.

NODE_DOWN_FORCED, *node*, *comments*

Stops cluster services on a specified node and places resource groups in an UNMANAGED state. Resources on the node remain online, that is they are *not* released.

node	The name of a node on which to stop cluster services
comments	User-defined text to describe the configured test.

Example
```
NODE_DOWN_FORCED, node2, Bring down node2 forced
```

Entrance Criteria
Cluster services on another node have *not* already been stopped with its resource groups placed in an UNMANAGED state. The specified node is active.

Success Indicators
The following conditions indicate success for this test:

- The cluster becomes stable
- The resource groups on the node change to UNMANAGED state
- Cluster services stop on the specified node
- Cluster services continue to run on other nodes
- All resource groups remain in the same state.

Network Tests for an IP Network

This section lists tests that bring network interfaces up or down on an IP network. The Cluster Test Tool requires two IP networks to run any of the tests described in this section. The second network provides network connectivity for the tool to run. The Cluster Test Tool verifies that two IP networks are configured before running the test.

NETWORK_UP_LOCAL, *node*, *network*, *comments*

Brings a specified network up on a specified node by running the **ifconfig up** command on the node.

node	The name of the node on which to run the **ifconfig up** command
network	The name of the network to which the interface is connected

comments	User-defined text to describe the configured test.

Example
```
NETWORK_UP_LOCAL, node6, hanet1, Bring up hanet1 on node 6
```

Entrance Criteria
The specified node is active and has at least one inactive interface on the specified network.

Success Indicators
The following conditions indicate success for this test:

- The cluster becomes stable
- Cluster services continue to run on the cluster nodes where they were active before the test
- Resource groups that are in the ERROR state on the specified node and that have a service IP label available on the network can go online, but should not enter the ERROR state
- Resource groups on other nodes remain in the same state.

NETWORK_DOWN_LOCAL, *node*, *network*, *comments*

Brings a specified network down on a specified node by running the **ifconfig down** command.

Note: If one IP network is already unavailable on a node, the cluster may become partitioned. The Cluster Test Tool does *not* take this into account when determining the success or failure of a test.

node	The name of the node on which to run the **ifconfig down** command
network	The name of the network to which the interface is connected
comments	User-defined text to describe the configured test.

Example
```
NETWORK_DOWN_LOCAL, node8, hanet2, Bring down hanet2 on node 8
```

Entrance Criteria
The specified node is active and has at least one active interface on the specified network.

Success Indicators
The following conditions indicate success for this test:

- The cluster becomes stable
- Cluster services continue to run on the cluster nodes where they were active before the test
- Resource groups on other nodes remain in the same state; however, some may be hosted on a different node
- If the node hosts a resource group for which the recovery method is set to notify, the resource group does *not* move.

NETWORK_UP_GLOBAL, *network, comments*

Brings specified network up on all nodes that have interfaces on the network. The network specified may be an IP network or a serial network.

network	The name of the network to which the interface is connected
comments	User-defined text to describe the configured test.

Example

```
NETWORK_UP_GLOBAL, hanet1, Bring up hanet1 on node 6
```

Entrance Criteria

Specified network is active on at least one node.

Success Indicators

The following conditions indicate success for this test:

- The cluster becomes stable
- Cluster services continue to run on the cluster nodes where they were active before the test
- Resource groups that are in the ERROR state on the specified node and that have a service IP label available on the network can go online, but should not enter the ERROR state
- Resource groups on other nodes remain in the same state.

NETWORK_DOWN_GLOBAL, *network, comments*

Brings the specified network down on all nodes that have interfaces on the network. The network specified may be an IP network or a serial network.

Note: If one IP network is already unavailable on a node, the cluster may become partitioned. The Cluster Test Tool does *not* take this into account when determining the success or failure of a test.

network	The name of the network to which the interface is connected
comments	User-defined text to describe the configured test.

Example

```
NETWORK_DOWN_GLOBAL, hanet1, Bring down hanet1 on node 6
```

Entrance Criteria

Specified network is inactive on at least one node.

Success Indicators

The following conditions indicate success for this test:

- The cluster becomes stable
- Cluster services continue to run on the cluster nodes where they were active before the test
- Resource groups on other nodes remain in the same state.

Network Interface Tests for IP Networks

JOIN_LABEL *iplabel, comments*

Brings up a network interface associated with the specified IP label on a specified node by running the **ifconfig up** command.

Note: You specify the IP label as the parameter. The interface that is currently hosting the IP label is used as the argument to the **ifconfig** command. The IP label can be a service, boot, or backup (standby) label. If it is a service label, then that service label must be hosted on some interface, for example, when the resource group is actually online. You can*not* specify a service label that is *not* already hosted on an interface.

The only time you could have a resource group online and the service label hosted on an inactive interface would be when the service interface fails but there was no place to move the resource group, in which case it stays online.

iplabel	The IP label of the interface.
comments	User-defined text to describe the configured test.

Example

```
JOIN_LABEL, app_serv_address, Bring up app_serv_address on node 2
```

Entrance Criteria

Specified interface is currently active on the specified node.

Success Indicators

The following conditions indicate success for this test:

- The cluster becomes stable
- Specified interface comes up on specified node
- Cluster services continue to run on the cluster nodes where they were active before the test
- Resource groups that are in the ERROR state on the specified node and that have a service IP label available on the network can go online, but should not enter the ERROR state
- Resource groups on other nodes remain in the same state.

FAIL_LABEL, *iplabel, comments*

Brings down a network interface associated with a specified label on a specified node by running the **ifconfig down** command.

Note: You specify the IP label as the parameter. The interface that is currently hosting the IP label is used as the argument to the **ifconfig** command The IP label can be a service, boot, or standby (backup) label.

iplabel	The IP label of the interface.
comments	User-defined text to describe the configured test.

Example

```
FAIL_LABEL, app_serv_label, Bring down app_serv_label, on node 2
```

Entrance Criteria

The specified interface is currently inactive on the specified node

Success Indicators

The following conditions indicate success for this test:

- The cluster becomes stable
- Any service labels that were hosted by the interface are recovered
- Resource groups that are in the ERROR state on the specified node and that have a service IP label available on the network can go online, but should not enter the ERROR state
- Resource groups remain in the same state; however, the resource group may be hosted by another node.

Network Tests for a Non-IP Network

The testing for non-IP networks is part of the NETWORK_UP_GLOBAL, NETWORK_DOWN_GLOBAL, NETWORK_UP_LOCAL and NETWORK_DOWN_LOCAL test procedures.

Resource Group Tests

RG_ONLINE, *rg, node* | ALL | ANY | RESTORE*, comments*

Brings a resource group online in a running cluster.

rg	The name of the resource group to bring online.
node	The name of the node where the resource group will come online.
ALL	Use **ALL** for concurrent resource groups only. When **ALL** is specified, the resource group will be brought online on all nodes in the resource group. If you use **ALL** for non-concurrent groups, the Test Tool interprets it as **ANY**.

ANY	Use **ANY** for non-concurrent resource groups to pick a node where the resource group is offline. For concurrent resource groups, use **ANY** to pick a random node where the resource group will be brought online.
RESTORE	Use **RESTORE** for non-concurrent resource groups to bring the resource groups online on the highest priority available node. For concurrent resource groups, the resource group will be brought online on all nodes in the nodelist.
comments	User-defined text to describe the configured test.

Example

```
RG_ONLINE, rg_1, node2, Bring rg_1 online on node 2.
```

Entrance Criteria

The specified resource group is offline, there are available resources, and can meet all dependencies.

Success Indicators

The following conditions indicate success for this test:

- The cluster becomes stable
- The resource group is brought online successfully on the specified node
- No resource groups go offline or into ERROR state.

RG_OFFLINE, *rg, node* | ALL | ANY, *comments*

Brings a resource group offline that is already online in a running cluster.

rg	The name of the resource group to bring offline
node	The name of the node on which the resource group will be taken offline
ALL	Use **ALL** for concurrent resource groups to bring the resource group offline on all nodes where the resource group is hosted. You can also use **ALL** for non-concurrent resource groups to bring the group offline on the node where it is online.
ANY	Use **ANY** for non-concurrent resource groups to bring the resource group offline on the node where it is online. You can use **ANY** for concurrent resource groups to select a random node where the resource group is online.
comments	User-defined text to describe the configured test

Example

```
RG_OFFLINE, rg_1, node2, Bring rg_1 offline from node2
```

Entrance Criteria

The specified resource group is online on the specified node

Success Indicators

The following conditions indicate success for this test:

- The cluster becomes stable
- Resource group, which was online on the specified node, is brought offline successfully
- Other resource groups remain in the same state.

RG_MOVE*, rg, node* | ANY | RESTORE*, comments*

Moves a resource group that is already online in a running cluster to a specific or any available node.

rg	The name of the resource group to bring offline
node	The target node; the name of the node to which the resource group will move
ANY	Use **ANY** to let the Cluster Test Tool pick a random available node to which to move the resource group.
RESTORE	Enable the resource group to move to the highest priority node available
comments	User-defined text to describe the configured test

Example

```
RG_MOVE, rg_1, ANY, Move rg_1 to any available node.
```

Entrance Criteria

The specified resource group must be non-concurrent and must be online on a node other *than* the target node.

Success Indicators

The following conditions indicate success for this test:

- The cluster becomes stable
- Resource group is moved to the target node successfully
- Other resource groups remain in the same state.

RG_MOVE_SITE, *rg, site* | OTHER, *comments*

Moves a resource group that is already online in a running cluster to an available node at a specific site.

rg	The name of the resource group to bring offline
site	The site where the resource group will move
OTHER	Use **OTHER** to have the Cluster Test Tool pick the "other" site as the resource group destination. For example, if the resource group is online on siteA, it will be moved to siteB, and conversely if the resource group is online on siteB, it will be moved to siteA.
comments	User-defined text to describe the configured test

Example

```
RG_MOVE_SITE, rg_1, site_2, Move rg_1 to site_2.
```

Entrance Criteria

The specified resource group is online on a node, *other* than the a node in the target site

Success Indicators

The following conditions indicate success for this test:

- The cluster becomes stable
- Resource group is moved to the target site successfully
- Other resource groups remain in the same state.

Volume Group Tests

VG_DOWN, *vg, node* | ALL | ANY, *comments*

Forces an error for a disk that contains a volume group in a resource group.

vg	The volume group on the disk of which to fail
node	The name of the node where the resource group that contains the specified volume group is currently online
ALL	Use **ALL** for concurrent resource groups. When **ALL** is specified, the Cluster Test Tool will fail the volume group on all nodes in the resource group where the resource group is online. If ALL is used for non-concurrent resource groups, the Tool performs this test for any resource group.

ANY	Use **ANY** to have the Cluster Test Tool will select the node as follows: • For a non-concurrent resource group, the Cluster Test Tool will select the node where the resource group is currently online. • For a concurrent resource group, the Cluster Test Tool will select a random node from the concurrent resource group node list, where the resource group is online
comments	User-defined text to describe the configured test.

Example

```
VG_DOWN, sharedvg, ANY, Fail the disk where sharedvg resides
```

Entrance Criteria

The resource group containing the specified volume groups is online on the specified node.

Success Indicators

The following conditions indicate success for this test:

- The cluster becomes stable
- Resource group containing the specified volume group successfully moves to another node, or if it is a concurrent resource groups, it goes into an ERROR state
- Resource groups may change state to meet dependencies.

Site Tests

SITE_ISOLATION, *comments*

Fails all the XD_data networks, causing the **site_isolation** event.

comments	User-defined text to describe the configured test.

Example

```
SITE_ISOLATION, Fail all the XD_data networks
```

Entrance Criteria

At least one XD_data network is configured and is up on any node in the cluster.

Success Indicators

The following conditions indicate success for this test:

- The XD_data network fails, no resource groups change state
- The cluster becomes stable.

SITE_MERGE, *comments*

Runs when at least one XD_data network is up to restore connections between the sites, and remove site isolation. Run this test after running the **SITE_ISOLATION** test.

comments	User-defined text to describe the configured test

Example

```
SITE_MERGE, Heal the XD_data networks
```

Entrance Criteria

At least one node must be online.

Success Indicators

The following conditions indicate success for this test:

- No resource groups change state
- The cluster becomes stable.

SITE_DOWN_TAKEOVER, *site, comments*

Stops cluster services and moves the resource groups to other nodes, on all nodes at the specified site.

site	The site that contains the nodes on which cluster services will be stopped
comments	User-defined text to describe the configured test

Example

```
SITE_DOWN_TAKEOVER, site_1, Stop cluster services on all nodes at
site_1, bringing the resource groups offline and moving the resource
groups.
```

Entrance Criteria

At least one node at the site must be online.

Success Indicators

The following conditions indicate success for this test:

- Cluster services are stopped on all nodes at the specified site
- All primary instance resource groups mover to the another site.
- All secondary instance resource groups go offline
- The cluster becomes stable.

SITE_UP, *site, comments*

Starts cluster services on all nodes at the specified site.

site	The site that contains the nodes on which cluster services will be started
comments	User-defined text to describe the configured test

Example

```
SITE_UP, site_1, Start cluster services on all nodes at site_1.
```

Entrance Criteria

At least one node at the site must be offline.

Success Indicators

The following conditions indicate success for this test:

- Cluster services are started on all nodes at the specified site
- Resource groups remain in the same state
- The cluster becomes stable.

General Tests

The other tests available to use in HACMP cluster testing:

- Bring an application server down
- Terminate the Cluster Manager on a node
- Add a wait time for test processing.

SERVER_DOWN, *node* | ANY, *appserv*, *command*, *comments*

Runs the specified command to stop an application server. This test is useful when testing application availability.

In the automated test, the test uses the stop script to turn off the application.

node	The name of a node on which the specified application sever is to become unavailable
ANY	Any available node that participates in this resource group can have the application server become unavailable The Cluster Test Tool tries to simulate server failure on any available cluster node. This test is equivalent to failure on the node that currently owns the resource group, *if the server is in a resource group that has policies other than the following ones*: • Startup: Online on all available nodes • Fallover: Bring offline (on error node only)
appserv	The name of the application server associated with the specified node
command	The command to be run to stop the application server
comments	User-defined text to describe the configured test.

Example

```
SERVER_DOWN,node1,db_app /apps/stop_db.pl, Kill the db app
```

Entrance Criteria

The resource group is online on the specified node.

Success Indicators

The following conditions indicate success for this test:

- The cluster becomes stable
- Cluster nodes remain in the same state

- The resource group that contains the application server is online; however, the resource group may be hosted by another node, unless it is a concurrent resource group, in which case the group goes into ERROR state.

CLSTRMGR_KILL, *node*, *comments*

Runs the **kill** command to terminate the Cluster Manager on a specified node.

Note: If **CLSTRMGR_KILL** is run on the local node, you may need to reboot the node. On startup, the Cluster Test Tool automatically starts again. For information about how to avoid manually rebooting the node, see the section Recovering the Control Node after Cluster Manager Stops.

For the Cluster Test Tool to accurately assess the success or failure of a **CLSTRMGR_KILL** test, do *not* perform other activities in the cluster while the Cluster Test Tool is running.

node	The name of the node on which to terminate the Cluster Manager
comments	User-defined text to describe the configured test.

Example

```
CLSTRMGR_KILL, node5, Bring down node5 hard
```

Entrance Criteria

The specified node is active.

Success Indicators

The following conditions indicate success for this test:

- The cluster becomes stable
- Cluster services stop on the specified node
- Cluster services continue to run on other nodes
- Resource groups that were online on the node where the Cluster Manager fails move to other nodes
- All resource groups on other nodes remain in the same state.

For information about potential conditions caused by a **CLSTRMGR_KILL** test running on the control node, see the section Recovering the Control Node after Cluster Manager Stops.

WAIT, *seconds*, *comments*

Generates a wait period for the Cluster Test Tool for a specified number of seconds.

seconds	The number of seconds that the Cluster Test Tool waits before proceeding with processing
comments	User-defined text to describe the configured test

Example
```
WAIT, 300, We need to wait for five minutes before the next test
```

Entrance Criteria
Not applicable.

Success Indicators
Not applicable.

Example Test Plan

The following excerpt from a sample Test Plan includes the tests:

- **NODE_UP**
- **NODE_DOWN_GRACEFUL**

It also includes a WAIT interval. The comment text at the end of the line describes the action to be taken by the test.

```
NODE_UP,ALL,starts cluster services on all nodes
NODE_DOWN_GRACEFUL,waltham,stops cluster services gracefully on node waltham
WAIT,20
NODE_UP,waltham,starts cluster services on node waltham
```

Running Custom Test Procedures

Before you start running custom tests, ensure that:

- Your Test Plan is configured correctly.

  For information about setting up a Test Plan, see the section Creating a Test Plan.
- You have specified values for test parameters.

  For information about parameter values, see the section Specifying Parameters for Tests.
- You have logging for the tool configured to capture the information that you want to examine for your cluster.

  For information about customizing verbose logging for the Cluster Test Tool, see the section Error Logging.
- The cluster is *not* in service in a production environment.

Launching a Custom Test Procedure

To run custom testing:

1. Enter `smit hacmp`
2. In SMIT, select either one of the following options:
   - **Extended Configuration**
   - **Problem Determination Tools**

   Then select **HACMP Cluster Test Tool.**
3. In the **HACMP Cluster Test Tool** panel, select **Execute Custom Test Procedure**.

4. In the **Execute Custom Test Procedure** panel, enter field values as follows:

Test Plan	(*Required*) The full path to the Test Plan for the Cluster Test Tool. This file specifies the tests for the tool to execute.
Variable File	*(Using a variables file is optional but recommended.*) The full path to the variables file for the Cluster Test Tool. This file specifies the variable definitions used in processing the Test Plan.
Verbose Logging	When set to **yes**, includes additional information in the log file that may help to judge the success or failure of some tests. For more information about verbose logging, see the section Running Automated Tests. The default is **yes**. Select **no** to decrease the amount of information logged by the Cluster Test Tool.
Cycle Log File	When set to **yes,** uses a new log file to store output from the Cluster Test Tool. The default is **yes**. Select **no** to append messages to the current log file. For more information about cycling the log file, see the section Log File Rotation.
Abort on Error	When set to **no,** the Cluster Test Tool continues to run tests after some of the tests being run fail. This may cause subsequent tests to fail because the cluster state is different from the one expected by one of those tests. The default is **no**. Select **yes** to stop processing after the first test fails. For information about the conditions under which the Cluster Test Tool stops running, see the section Cluster Test Tool Stops Running. **Note:** The tool stops running and issues an error if a test fails and **Abort on Error** is selected.

5. Press Enter to start running the custom tests.
6. Evaluate the test results.

   For information about evaluating test results, see the section Evaluating Results.

Evaluating Results

You evaluate test results by reviewing the contents of the log file created by the Cluster Test Tool. When you run the Cluster Test Tool from SMIT, it displays status messages to the screen and stores output from the tests in the file **/var/hacmp/log/cl_testtool.log**. Messages indicate when a test starts and finishes and provide additional status information. More detailed information, especially when verbose logging is enabled, is stored in the log file that appears on the screen. Information is also logged to the **hacmp.out** file. For information about the **hacmp.out** file, see Chapter 2: Using Cluster Log Files in the *Troubleshooting Guide*.

Criteria for Test Success or Failure

The following criteria determine the success or failure of cluster tests:

- Did the cluster stabilize?

  For the Cluster Test Tool, a cluster is considered stable when:

  - The Cluster Manager has a status of stable on each node, or is not running.
  - Nodes that should be online are online.

    If a node is stopped and that node is the last node in the cluster, the cluster is considered stable when the Cluster Manager is inoperative on all nodes.
  - No events are in the event queue for HACMP.

  The Cluster Test Tool also monitors HACMP timers that may be active. The tool waits for some of these timers to complete before determining cluster stability. For more information about how the Cluster Test Tool interacts with HACMP timers, see the section Working with Timer Settings.
- Has an appropriate recovery event for the test run?
- Is a specific node online or offline as specified?
- Are all expected resource groups still online within the cluster?
- Did a test that was expected to run actually run?

Every test checks to see if it makes sense to be run; this is called a check for “rationality”. A test returning a `NOT RATIONAL` status indicates the test could *not* be run because the entrance criteria could *not* be met; for example, trying to run the NODE_UP test on a node that is already up. A warning message will be issued along with the exit status to explain why the test was *not* run. Irrational tests do not cause the Cluster Test Tool to abort.

The `NOT RATIONAL` status indicates the test was *not* appropriate for your cluster. When performing automated testing, it is important to understand why the test did *not* run. For Custom Cluster tests, check the sequences of events and modify the test plan to ensure the test runs. Consider the order of the tests and the state of the cluster before running the test plan. For more information, refer to the section Setting up Custom Cluster Testing.

The tool targets availability as being of primary importance when reporting success or failure for a test. For example, if the resource groups that are expected to be available are available, the test passes.

Keep in mind that the Cluster Test Tool is testing the cluster configuration, *not* testing HACMP. In some cases the configuration may generate an error that causes a test to fail, even though the error is the expected behavior. For example, if a resource group enters the error state and there is no node to acquire the resource group, the test fails.

Note: If a test generates an error, the Cluster Test Tool interprets the error as a test failure. For information about how the Cluster Test Tool determines the success or failure of a test, see the Success Indicators subsections for each test in the section Description of Tests.

Recovering the Control Node after Cluster Manager Stops

If a **CLSTRMGR_KILL** test runs on the control node and stops the control node, reboot the control node. No action is taken to recover from the failure. After the node reboots, the testing continues.

To monitor testing after the Cluster Test Tool starts again, review output in the **/var/hacmp/log/cl_testtool.log** file. To determine whether a test procedure completes, run the **tail -f** command on **/var/hacmp/log/cl_testtool.log** file.

How to Avoid Manual Intervention

You can avoid manual intervention to reboot the control node during testing by:

- Editing the **/etc/cluster/hacmp.term** file to change the default action after an abnormal exit.

  The **clexit.rc** script checks for the presence of this file and, if the file is executable, the script calls it instead of halting the system automatically.
- Configuring the node to auto-Initial Program Load (IPL) before running the Cluster Test Tool.

Error Logging

The Cluster Test Tool has several useful functions that enable you to work with logs.

Log Files: Overview

If a test fails, the Cluster Test Tool collects information in the automatically created log files.To collect logs, the Cluster Test Tool creates the directory **/var/hacmp/cl_testtool** if it doesn't exist. HACMP never deletes the files in this directory.You evaluate the success or failure of tests by reviewing the contents of the Cluster Test Tool log file, **/var/hacmp/utilities/cl_testtool.log**.

For each test plan that has any failures, the tool creates a new directory under **/var/hacmp/cl_testtool**. If the test plan has no failures, the tool does *not* create a log directory. The directory name is unique and consists of the name of the Cluster Test Tool plan file, and the time stamp when the test plan was run.

Log File Rotation

The Cluster Test Tool saves up to three log files and numbers them so that you can compare the results of different cluster tests. The tool also rotates the files with the oldest file being overwritten. The following list shows the three files saved:

/var/hacmp/utilities/cl_testtool.log

/var/hacmp/utilities/cl_testtool.log.1

/var/hacmp/utilities/cl_testtool.log.2

If you do *not* want the tool to rotate the log files, you can disable this feature from SMIT. For information about turning off this feature, see the section Running Automated Tests or Setting up Custom Cluster Testing.

Log File Entries

The entries in the log file are in the format:

```
DD/MM/YYYY_hh:mm:ss Message text . . .
```

where `DD/MM/YYYY_hh:mm:ss` indicates day/month/year_hour/minutes/seconds.

The following example shows the type of output stored in the log file:

```
04/02/2006/_13:21:55:
-------------------------------------------------------
04/02/2006/_13:21:55: | Initializing Variable Table
04/02/2006/_13:21:55:
-------------------------------------------------------
04/02/2006/_13:21:55:   Using Variable File: /tmp/sample_variables
04/02/2006/_13:21:55:           data line: node1=waltham
04/02/2006/_13:21:55:           key: node1 - val: waltham
04/02/2006/_13:21:55:
-------------------------------------------------------
04/02/2006/_13:21:55: | Reading Static Configuration Data
04/02/2006/_13:21:55:
-------------------------------------------------------
04/02/2006/_13:21:55:   Cluster Name: Test_Cluster
04/02/2006/_13:21:55:   Cluster Version: 7
04/02/2006/_13:21:55:   Local Node Name: waltham
04/02/2006/_13:21:55:   Cluster Nodes: waltham belmont
04/02/2006/_13:21:55:   Found 1 Cluster Networks
04/02/2006/_13:21:55:   Found 4 Cluster Interfaces/Device/Labels
04/02/2006/_13:21:55:   Found 0 Cluster Resource Groups
04/02/2006/_13:21:55:   Found 0 Cluster Resources
04/02/2006/_13:21:55:   Event Timeout Value: 720
04/02/2006/_13:21:55:   Maximum Timeout Value: 2880
04/02/2006/_13:21:55:
-------------------------------------------------------
04/02/2006/_13:21:55: | Building Test Queue
04/02/2006/_13:21:55:
-------------------------------------------------------
04/02/2006/_13:21:55:   Test Plan: /tmp/sample_event
04/02/2006/_13:21:55:           Event 1: NODE_UP: NODE_UP,ALL,starts
cluster services on all nodes
04/02/2006/_13:21:55:
-------------------------------------------------------
04/02/2006/_13:21:55: | Validate NODE_UP
04/02/2006/_13:21:55:
-------------------------------------------------------
04/02/2006/_13:21:55:   Event node: ALL
04/02/2006/_13:21:55:   Configured nodes: waltham belmont
04/02/2006/_13:21:55:           Event 2: NODE_DOWN_GRACEFUL:
NODE_DOWN_GRACEFUL,node1,stops cluster services gracefully on node1
04/02/2006/_13:21:55:
-------------------------------------------------------
04/02/2006/_13:21:55: | Validate NODE_DOWN_GRACEFUL
04/02/2006/_13:21:55:
-------------------------------------------------------
04/02/2006/_13:21:55:   Event node: waltham
04/02/2006/_13:21:55:   Configured nodes: waltham belmont
04/02/2006/_13:21:55:           Event 3: WAIT: WAIT,20
04/02/2006/_13:21:55:           Event 4: NODE_UP: NODE_UP,node1,starts
cluster services on node1
```

```
04/02/2006/_13:21:55:
-------------------------------------------------------
04/02/2006/_13:21:55: | Validate NODE_UP
04/02/2006/_13:21:55:
-------------------------------------------------------
04/02/2006/_13:21:55:   Event node: waltham
04/02/2006/_13:21:55:   Configured nodes: waltham belmont
04/02/2006/_13:21:55:
.
.
.
```

Log File Example

If a test fails, you will see output similar to the following:

```
======================================================================
Test 1 Complete - NETWORK_DOWN_LOCAL: fail service network

Test Completion Status: FAILED

======================================================================

Copying log files hacmp.out and clstrmgr.debug from all nodes to
directory /var/hacmp/cl_testtool/rg_fallover_plan.1144942311
on node prodnode1.
```

After that, you can examine the directory **/var/hacmp/cl_testtool/rg_fallover_plan.1144942311** on node **prodnode1**.

In the log directory, the tool creates separate files for each test. The names for the specific log files stored in the directory have this structure:

`<testnum>.<testname>.<node>.<logfile>`

where

- `testnum` is the order in which the test appears in the test plan file
- `testname` is the name of the test that failed
- `node` is the node from which the log was collected
- `logfile` the source of the logging information, either the **hacmp.out** or **clstrmgr.debug** files.

For example, if the NETWORK_DOWN_LOCAL test fails and it is the first test that was run, and later in the test plan the fourth test, named RG_MOVE also fails, you will see the following files in the **/var/hacmp/cl_testtool/rg_fallover_plan.1144942311** directory:

```
1.NETWORK_DOWN_LOCAL.prodnode1.clstrmgr.debug
1.NETWORK_DOWN_LOCAL.prodnode1.hacmp.out
1.NETWORK_DOWN_LOCAL.prodnode2.clstrmgr.debug
1.NETWORK_DOWN_LOCAL.prodnode2.hacmp.out
4.RG_MOVE.prodnode1.clstrmgr.debug
4.RG_MOVE.prodnode1.hacmp.out
4.RG_MOVE.prodnode2.clstrmgr.debug
4.RG_MOVE.prodnode2.hacmp.out
```

The hacmp.out File

The **hacmp.out** file also logs the start of each test that the Cluster Test Tool runs on each cluster node. This log entry has the following format:

TestName: *datetimestring1*:*datetimestring2*

where

TestName	The name of the test being processed.
datetimestring1	The date and time on the control node when the Cluster Test Tool starts to run the test. The value of *datetimestring* has the format MMDDHHmmYY (month day hour minute year).
datetimestring2	The date and time on the node on which the test runs. The value of *datetimestring* has the format MMDDHHmmYY (month day hour minute year).

Note: The Cluster Test Tool uses the date and time strings to query the AIX 5L error log when necessary.

Verbose Logging: Overview

By default, the Cluster Test Tool uses verbose logging to provide a wealth of information about the results of cluster testing. You can customize the type of information that the tool gathers and stores in the Cluster Test Tool log file.

Note: The Cluster Snapshot utility does *not* include the Cluster Test Tool log file because this file is specific to HACMP cluster testing at a specific point in time—*not* an indication of *ongoing* cluster status.

With verbose logging enabled, the Cluster Test Tool:

- Provides detailed information for each test run
- Runs the following utilities on the control node between the processing of one test and the next test in the list:

Utility	Type of Information Collected
clRGinfo	The location and status of resource groups
errpt	Errors stored in the system error log file

- Processes each line in the following files to identify additional information to be included in the Cluster Test Tool log file. The utilities included are run on each node in the cluster after a test finishes running.

File	Type of Information Specified
cl_testtool_log_cmds	A list of utilities to be run to collect additional status information See the section Customizing the Types of Information to Collect.
cl_testtool_search_strings	Text strings that may be in the **hacmp.out** file. The Cluster Test Tool searches for these strings and inserts any lines that match into the Cluster Test Tool log file. See the section Adding Data from hacmp.out to the Cluster Test Tool Log File.

If you want to gather only basic information about the results of cluster testing, you can disable verbose logging for the tool. For information about disabling verbose logging for the Cluster Test Tool, see the section Running Automated Tests or Setting up Custom Cluster Testing.

Customizing the Types of Information to Collect

You can customize the types of logging information to be gathered during testing. When verbose logging is enabled for the Cluster Test Tool, it runs the utilities listed in the **/usr/es/sbin/cluster/etc/cl_testtool_log_cmds** file, and collects status information that the specified commands generate. The Cluster Test Tool runs each of the commands listed in **cl_testtool_log_cmds** file after each test completes, gathers output for each node in the cluster, and stores this information in the Cluster Test Tool log file.

You can collect information specific to a node by adding or removing utilities from the list. For example, if you have an application server running on two of the nodes in a four-node cluster, you could add application-specific commands to the list on the nodes running the application servers.

If you want all of the cluster nodes to use the same **cl_testtool_log_cmds** file, you can add it to a file collection. For information about including files in a file collection, see Chapter 7: Verifying and Synchronizing an HACMP Cluster.

By default, the **cl_testtool_log_cmds** file includes the following utilities:

Utility	Type of Information Collected
/usr/es/sbin/cluster/utilities/cldump	A snapshot of the status of key cluster components—the cluster itself, the nodes in the cluster, the network interfaces connected to the nodes, and the resource groups on each node
lssrc -ls clstrmgrES	The status of the Cluster Manager
lssrc -ls topsvcs	The status of Topology Services

The file also contains entries for the following utilities, but they are commented out and *not* run. If you want to run any of these utilities between each test, open the file and remove the comment character from the beginning of the command line for the utility.

Utility	Type of Information Collected
snmpinfo -m dump -v -o /usr/es/sbin/cluster/hacmp.defs cluster	Information on MIB cluster status
snmpinfo -m dump -v -o /usr/sbin/cluster/hacmp.defs resGroupNodeState	Information on MIB resource group state
LANG=C lssrc -a | grep -vw “inoperative$”	The status of all subsystems for each host
svmon -C clstrmgr	Memory usage statistics for the Cluster Manager
/usr/sbin/rsct/bin/hatsdmsinfo	Information about the deadman switch timer
netstat -i ; netstat -r	Information about configured interfaces and routes
lssrc -ls gsclvmd	Information about **gsclvmd**—the access daemon for enhanced concurrent mode volume groups
ps auxw	Process information
lsvg -o	Information about active volume groups (those that are varied on and accessible)
lspv	Information about the physical volumes in a volume group
vmstat; vmstat -s	System resource utilization information that includes statistics for virtual memory, kernel, disks, traps, and CPU activity

You can also add and remove commands from the **cl_testtool_log_cmds** file.

Note: Enter only one command on each line of the file. The tool executes one command per line.

Adding Data from hacmp.out to the Cluster Test Tool Log File

You can add messages that include specified text in the **hacmp.out** file to the Cluster Test Tool log file. With verbose logging enabled, the tool uses the **/usr/es/sbin/cluster/etc/cl_testtool/cl_testtool_search_strings** file to identify text strings to search for in **hacmp.out**. For any text string that you specify on a separate line in the **cl_testtool_search_strings** file, the tool:

- Searches the **hacmp.out** file for a matching string
- Logs the line containing that string, accompanied by the line number from the **hacmp.out** file, to the Cluster Test Tool log file

You can use the line number to locate the line in the **hacmp.out** file and then review that line within the context of other messages in the file.

By default, the file contains the following lines:

```
!!!!!!!!!!   ERROR   !!!!!!!!!!
          EVENT FAILED
```

You can edit the **cl_testtool_search_strings** file on each node to specify a search string specific to a node. This way, the **cl_testtool_search_strings** file is different on different nodes.

If you want all of the cluster nodes to use the same **cl_testtool_search_strings** file, you can add it to a file collection and synchronize the cluster. For information about including files in a file collection, see Chapter 7: Verifying and Synchronizing an HACMP Cluster.

Note: Cluster synchronization does *not* propagate a **cl_testtool_search_strings** file to other nodes in a cluster unless the file is part of a file collection.

To edit the **cl_testtool_search_strings** file:

- On each line of the file, specify a single text string that you want the tool to locate in the **hacmp.out** file.

Fixing Problems when Running Cluster Tests

This section discusses the following issues that you may encounter when testing a cluster:

- Cluster Test Tool Stops Running
- Control Node Becomes Unavailable
- Cluster Does Not Return to a Stable State
- Working with Timer Settings
- Testing Does Not Progress as Expected
- Unexpected Test Results.

Cluster Test Tool Stops Running

The Cluster Test Tool can stop running under the following conditions:

- The Cluster Test Tool fails to initialize
- A test fails and **Abort on Error** is set to **yes** for the test procedure
- The tool times out waiting for cluster stabilization, or the cluster fails to stabilize after a test.

  See the section Working with Timer Settings
- An error that prohibits the Cluster Test Tool from running a test, such as a configuration in AIX 5L or a script that is missing
- A cluster recovery event fails and requires user intervention.

Control Node Becomes Unavailable

If the control node experiences an unexpected failure while the Cluster Test Tool is running, the testing stops. No action is taken to recover from the failure.

To recover from the failure:

1. Bring the node back online and start cluster services in the usual manner.

   You may need to reboot the control node.
2. Stabilize the cluster.
3. Run the test again.

Note: The failure of the control node may invalidate the testing that occurred prior to the failure.

If a **CLSTRMGR_KILL** test runs on the control node, the node and cluster services need to restart. For information about handling this situation, see the section Recovering the Control Node after Cluster Manager Stops.

Cluster Does Not Return to a Stable State

The Cluster Test Tool stops running tests after a timeout if the cluster does *not* return to a stable state either:

- While a test is running
- As a result of a test being processed.

The timeout is based on ongoing cluster activity and the cluster-wide event-duration time until warning values. If the Cluster Test Tool stops running, an error appears on the screen and is logged to the Cluster Test Tool log file before the tool stops running.

After the cluster returns to a stable state, it is possible that the cluster components, such as resource groups, networks, and nodes, are *not* in a state consistent with the specifications of the list of tests. If the tool can*not* run a test due to the state of the cluster, the tool generates an error. The Cluster Test Tool continues to process tests.

If the cluster state does *not* let you continue a test, you can:

1. Reboot cluster nodes and restart the Cluster Manager.
2. Inspect the Cluster Test Tool log file and the **hacmp.out** file to get more information about what may have happened when the test stopped.
3. Review the timer settings for the following cluster timers, and make sure that the settings are appropriate to your cluster:
   - Time until warning
   - Stabilization interval
   - Monitor interval.

   For information about timers in the Cluster Test tool, and about how application monitor timers can affect whether the tool times out, see the section Working with Timer Settings.

Working with Timer Settings

The Cluster Test Tool requires a stable HACMP cluster for testing. If the cluster becomes unstable, the time that the tool waits for the cluster to stabilize depends on the activity in the cluster:

- No activity.

  The tool waits for twice the time until event duration time until warning (also referred to as **config_too_long)** interval, then times out.
- Activity present.

  The tool calculates a timeout value based on the number of nodes in the cluster and the setting for the time until warning interval.

If the time until warning interval is too short for your cluster, testing may time out. To review or change the setting for the time until warning interval, in HACMP SMIT, select **HACMP Extended Configuration > Extended Performance Tuning Parameters Configuration** and press Enter.

For complete information on tuning event duration time, see the section Tuning Event Duration Time Until Warning in Chapter 5: Configuring Cluster Events.

The settings for the following timers configured for an application monitor can also affect whether testing times out:

- Stabilization interval
- Monitor interval

The settling time for resource groups does *not* affect whether or *not* the tool times out.

Stabilization Interval for an Application Monitor

If this timer is active, the Cluster Test Tool does *not* time out when waiting for cluster stability. If the monitor fails, however, and recovery actions are underway, the Cluster Test Tool may time out before the cluster stabilizes.

Make sure the stabilization interval configured in HACMP is appropriate for the application being monitored.

For information about setting the stabilization interval for an application, see Chapter 4: Configuring HACMP Cluster Topology and Resources (Extended).

Monitor Interval for a Custom Application Monitor

When the Cluster Test Tool runs a **server_down** test, it waits for the length of time specified by the monitor interval before the tool checks for cluster stability. The monitor interval defines how often to poll the application to make sure that the application is running.

The monitor interval should be long enough to allow recovery from a failure. If the monitor interval is too short, the Cluster Test Tool may time out when a recovery is in process.

For information about setting the monitor interval for an application, see Chapter 4: Configuring HACMP Cluster Topology and Resources (Extended).

Testing Does Not Progress as Expected

If the Cluster Test Tool is *not* processing tests and recording results as expected, use the Cluster Test Tool log file to try to resolve the problem:

1. Ensure that verbose logging for the tool is enabled.

For information about verbose logging for the Cluster test Tool, see the section Error Logging.

2. View logging information from the Cluster Test Tool log file **/var/hacmp/utilities/cl_testtool.log**. The tool directs more information to the log file than to the screen.
3. Add other tools to the **cl_testtool_log_cmds** file to gather additional debugging information. This way you can view this information within the context of the larger log file.

   For information about adding commands to the **cl_testtool_log_cmds** file, see the section Customizing the Types of Information to Collect.

Unexpected Test Results

The basic measure of success for a test is availability. In some instances, you may consider that a test has passed, when the tool indicates that the test failed. Be sure that you are familiar with the criteria that determines whether a test passes or fails. For information about the criteria for a test passing or failing, see the section Evaluating Results.

Also ensure that:

- Settings for cluster timers are appropriate to your cluster. See the section Cluster Does Not Return to a Stable State.
- Verbose logging is enabled and customized to investigate an issue. See the section Testing Does Not Progress as Expected.

Chapter 9: Starting and Stopping Cluster Services

This chapter explains how to start and stop cluster services on cluster nodes and clients. The following sections describe these options in detail:

Overview

HACMP 5.4 includes new features when you start or stop cluster services:

- *Start cluster services*. When you start the cluster services, HACMP by default automatically activates the resources according to how you defined them, taking into consideration application dependencies, application start and stop scripts, dynamic attributes and other parameters. That is, HACMP automatically manages (and activates, if needed) resource groups and applications in them.

  Note that you can also start HACMP with the option to manage resource groups manually. This tells HACMP *not to acquire any resource groups* (and applications) automatically for you.

  With HACMP 5.4, you can start HACMP cluster services on the node(s) without stopping your applications, by selecting an option from SMIT (**System Management (C-SPOC) > Manage HACMP Services > Start Cluster Services**).

  HACMP relies on the application monitor and application startup script to verify whether it needs to start the application for you or whether the application is already running (HACMP attempts *not* to start a second instance of the application).
- *Shut down the cluster services*. During an HACMP shutdown, you may select one of the following three actions for the resource groups:
  - Bring Resource Groups Offline.
  - Move Resource Groups to other node(s).
  - Unmanage Resource Groups.

  For more information on resource group states, see Appendix B: Resource Group Behavior during Cluster Events.

Starting Cluster Services

In HACMP 5.4, you can allow your applications that run outside of HACMP to continue running during installation of HACMP and when starting HACMP. There is no need to stop, restart or reboot the system or applications.

A Note on Application Monitors

HACMP 5.4 checks for running applications by using the configured application monitor. If the monitor indicates that the application is already running, HACMP will *not* start the second instance of the application. If the application monitors are not configured to HACMP, then you may write an application start script that checks the state of the application before starting it.

Application monitors, configurable in HACMP, are a critical piece of the HACMP cluster configuration; they enable HACMP to keep applications highly available. When HACMP starts an application server on a node, it also periodically monitors the application (using the monitor that you configure) to make sure that the application is up and running.

An erroneous application monitor may *not* detect a failed application. As a result, HACMP would *not* recover it or may erroneously detect an application as failed, which may cause HACMP to move the application to a takeover node, resulting in unnecessary downtime. To summarize, we highly recommend properly configured and tested application monitors for all applications that you want to keep highly available with the use of HACMP. Use them as follows:

- Use a process monitor if the intent is to monitor whether the process(es) exist on the UNIX system.
- Use a custom monitor if the intent is to check the health of the application, for example, whether the database is still functioning by querying a database table.
- Use both process and custom monitors when needed.

During verification, HACMP issues a warning if an application monitor is *not* configured.

For information on configuring an application monitor, see Configuring Multiple Application Monitors in Chapter 4: Configuring HACMP Cluster Topology and Resources (Extended).

Procedure for Starting Cluster Services

To start HACMP cluster services, as the root user, perform the following steps:

Note: Perform the following only after configuring and synchronizing the cluster. For more information, see Chapter 3: Configuring an HACMP Cluster (Standard).

1. Enter the fastpath `smit cl_admin`
2. In SMIT, select **Manage HACMP Services** > **Start Cluster Services** and press Enter.

3. Enter field values as follows:

Start now, on system restart or both	Indicate how you want to start cluster services when you commit the values on this panel by pressing Enter (**now**), when the operating system reboots by selecting **on system restart**, or on **both** occasions. Choosing **on system restart** or **both** means that the cluster services are always brought up automatically after a system reboot. **Note**: When you start the HACMP cluster services with the **Manage Resource Group** option set to **Manually**, and select the option **both**, the timing of a power loss or rebooting the node may affect whether the node is in the OFFLINE or UNMANAGED state after the system reboot.
Start Cluster Services on these nodes	Enter the name(s) of one or more nodes on which you want to start cluster services. Alternatively, you can select nodes from a picklist. Separate multiple nodes with a comma.
Manage Resource Groups	**Automatically** (default). HACMP brings resource group(s) online according to the resource groups' configuration settings and the current cluster state and starts managing the resource group(s) and applications for availability. When you start HACMP cluster services and set the Manage Resource Group option to Automatically, HACMP automatically activates resource groups on the node(s) according to their policies and locations and also starts applications. HACMP may not necessarily start the application on the same node on which it is currently being run, if the application is already running. That is, when this option is selected, HACMP determines on which node to bring a resource group online based on the configured resource group policies, resource group dependency configuration and available resources on the node. If you select this option while starting the cluster services, it is suggested to stop the applications and resources so that HACMP can start them on the appropriate node.

	Manually. HACMP does *not* activate resource groups while the cluster services on the selected node are started. After you start cluster services, you can bring any resource groups online or offline, as needed, using the **HACMP Resource Group and Application Management** SMIT menu (**clRGmove**). For more information, see Starting HACMP Cluster Services with Manually Managed Resource Groups.
BROADCAST message at startup?	Indicate whether you want to send a broadcast message to all nodes when the cluster services start. The default is **true**. Alternately, you can set the broadcast message setting for each local node using the **Extended Configuration > Extended Cluster Service Settings** path.
Startup Cluster Information Daemon?	Indicate whether you want to start the **clinfoES** daemon. For example, if your application uses the Cluster Information daemon, if you use the **clstat** monitor, or you want to run event emulation, set this field to **true.** Otherwise, set it to **false**. The value that you enter in the **Startup Cluster Information Services?** field works in conjunction with the value you enter in the **Start now, on system restart or both** field. If you set either (or both) of the startup fields to **true** and the **Start now, on system restart or both** field to **both**, then the **clinfoES** daemon is also started whenever the cluster services are started.
Ignore Verification Errors?	Set this value to false (the default) to stop all selected nodes from starting cluster services if verification finds errors on any node. Set this value to **true** to start cluster services even if verification finds errors on the specified nodes or in the cluster in general. This setting should be used with caution. For more information, see the section Automatic Verification and Synchronization.

Automatically correct errors found during cluster start?

Note: This field is available only if the automatic verification and synchronization option has been enabled. For more information, see Modifying the Startup of Cluster Services.

- Select **Interactively** to receive prompts to correct certain errors as they are found during verification. (If you are using WebSMIT, the **Interactively** option is *not* available in WebSMIT.)
- Select **No** if you do *not* want HACMP to correct any verification errors automatically. If you select **No**, you must correct errors, if any, manually.
- Select **Yes** if you want HACMP to correct cluster verification errors automatically without first prompting you.

  Note: *Not* all verification errors are automatically corrected; some must be corrected manually. For more information, see the section Automatic Verification and Synchronization.

4. Press Enter.

   The system performs verification and synchronization as needed, and then starts the cluster services on the nodes specified, activating the cluster configuration that you have defined. The time that it takes the commands and scripts to run depends on your configuration (for example, the number of disks, the number of interfaces to configure, the number of filesystems to mount, and the number of applications being started).

   SMIT displays a command status window. Note that when the SMIT panel indicates the completion of the cluster startup, HACMP processing of the resource groups in most cases has *not* yet completed. To verify that the processing has completed, use **/usr/es/sbin/cluster/clstat**, described in *Chapter 10: Monitoring an HACMP Cluster*.

Starting HACMP Cluster Services with Manually Managed Resource Groups

Set the cluster services **Manage Resource Group** startup option to **Manually** when you want more control over the node on which an application should run. This method ensures that the services provided by the application server are not interrupted.

When you choose this option to start the HACMP cluster services, the resource groups on the node remain in the OFFLINE or UNMANAGED state, depending on whether this is a cold start up or a start after the node was stopped and resource groups placed in an UNMANAGED state.

Note: Please be advised that if a resource group is in the UNMANAGED state, it does *not* mean that, from HACMP's point of view, the actual resources in the resource group are *not* running. To HACMP, it means that HACMP is *not* managing the resources (and the applications) of the resource group for availability.

Note that either you must have an application monitor configured that HACMP uses to check the application or your application start scripts should be intelligent enough not to start the application if it is already running.

If you want to activate resource groups that are *not* brought online automatically, use the Resource Group Management utility (**clRGmove**) to bring the OFFLINE state resource groups to the ONLINE state.

Consider the following example: If an application is running on a node that is *not* the primary node, and during the startup process you know that HACMP will move the resource group with the application to another node (according to the resource group policy specified), starting HACMP cluster services with the **Manage Resource Group** option set to **Manually** tells HACMP *not* to start the resource groups during startup. You can later use the user-requested **rg-move** to bring the resource group to the ONLINE state on the same node where your application is already running.

To start cluster services on a resource group that is manually managed:

1. Enter `smitty hacmp`
2. **System Management (C-SPOC) > Manage HACMP Services > Bring Resource Group Online**.
3. Select the node where your application is running.
4. Press Enter.

Starting Cluster Services on a Node with a Resource Group in the UNMANAGED State

Resource groups may be in the UNMANAGED state on a node if cluster services on that node have been stopped using the **Unmanage Resource Groups** option. (For more information, see Stopping Cluster Services later in this chapter.)

This **Unmanage Resource Groups** option causes HACMP to stop providing high availability services to the resource group; that is, the resource groups will not fall over due to resource failures. This option is intended for temporary situations, such as when you want to upgrade HACMP or perform maintenance without bringing your applications offline.

Starting cluster services on the node after it had been stopped with the resource group option set to UNMANAGED, therefore, puts any resource group that is in the UNMANAGED state on that node back to the state in which it was prior to being UNMANAGED. While bringing the resource group ONLINE from the UNMANAGED state, HACMP checks every resource in the resource group to see whether it is active and activates it if it is found inactive. Thus, it is critical to configure the application monitors so that HACMP can correctly detect a running application and so HACMP does *not* try to start a second instance.

In cases where you want to bring a resource group from an UNMANAGED state to an ONLINE state on a different node (because the node that was stopped using UNMANAGED option is unavailable), you should do the following:

1. Bring the resource groups to the OFFLINE state using a user-requested **rg-move** SMIT panel. Note that during this operation, HACMP will not stop any resources as the node that originally hosted the resource group is no longer available.
2. Ensure that all the resources that are configured in the resource group are OFFLINE, including the application, if any.

3. Bring the resource groups from their OFFLINE state to the ONLINE state, just as was necessary in previous releases using the resource group migration utility **clRGmove** or the SMIT option.

   Note: In the case where the node that was stopped is still available, moving a resource group from an UNMANAGED state to OFFLINE state would result in the stopped node actually releasing the resources of the resource group.

Modifying the Startup of Cluster Services

Typically, you should use the default cluster services startup settings—especially the verification setting, which is automatically enabled to ensure a safe startup. However, you can modify these settings by following the procedure described below.

Procedure for Modifying Startup of Cluster Services

To modify the startup of cluster services:

1. Enter the fastpath `smit cl_admin` or `smitty hacmp`.
2. Select **Extended Configuration** > **Extended Cluster Service Settings** and press Enter.
3. Enter field values in the SMIT panel as follows:

Start HACMP at system restart	**False** is the default. This removes the entry from the **/etc/inittab** file and will *not* automatically start cluster services at system restart. **True** starts the daemons after a system reboot by adding an entry to the **/etc/inittab** file.
BROADCAST message at startup	**True** is the default. This broadcasts a message to the console, indicating that cluster services are starting.
Startup Cluster Information Daemon?	**False** is the default. **True** starts the **clinfo** daemon, which allows **clstat** and **xclstat** (or any third-party application written against the **clinfo** API) to read changes in the cluster state.
Verify Cluster Prior to Startup?	**True** is the default. This ensures that HACMP will automatically verify and synchronize your cluster configuration before starting the cluster services. It is recommended that this value be set to **True.** Setting this value to **False** disables verification and synchronization from automatically occurring before the startup of cluster services.

Stopping Cluster Services

You typically stop cluster services:

- Before making any hardware or software changes or other scheduled node shutdowns or reboots. Failing to do so may cause unintended cluster events to be triggered on other nodes.
- Before certain reconfiguration activity. Some changes to the cluster information stored in the Configuration Database require stopping and restarting the cluster services on *all* nodes for the changes to become active. For example, if you wish to change the name of the cluster, the name of a node, or the name of a network interface, you must stop and restart cluster services on that node or on all nodes, depending on the cluster setup.

  For more information about which changes to the cluster require HACMP reconfiguration, see Appendix A: 7x24 Maintenance.

When stopping cluster services, minimize activity on the system. If the node you are stopping is currently providing highly available services, notify users of your intentions if their applications will be unavailable. Let them know when services will be restored.

Procedure for Stopping Cluster Services

The steps below describe the procedure for stopping cluster services on a single node or on all nodes in a cluster by using the C-SPOC utility on one of the cluster nodes.

To stop cluster services:

1. Enter the fastpath `smit cl_admin or smitty hacmp.`
2. Select **System Management (C-SPOC)** and press Enter.
3. Select **Manage HACMP Services > Stop Cluster Services** and press Enter.
4. Enter field values in the SMIT panel as follows:

Select an Action on Resource Groups	Indicate the type of shutdown: • **Bring Resource Groups Offline**: HACMP stops all managed resources currently ONLINE on the node being stopped. HACMP will *not* activate these resources on any other nodes, that is, no fallover. This option is equivalent to the option to stopping cluster services gracefully in previous releases. After successfully stopping all managed resources, HACMP stops RSCT services and goes into ST_INIT state.

- **Move Resource Groups**. HACMP stops all managed resources currently ONLINE on the node being stopped. The resource groups will be moved to a takeover node according to the configured resource group policies (if defined), dependency configurations (if defined) and available resources.

  This option is equivalent to the graceful with takeover option in previous releases.

  After successfully stopping all managed resources HACMP, stops RSCT services and the Cluster Manager daemon goes into ST_INIT state.

- **Unmanage Resource Groups**. The cluster services are stopped immediately. Resources that are online on the node are not stopped. Applications continue to run. This option is equivalent to the forced down option in previous releases.

  For more information, see Stopping HACMP Cluster Services without Stopping Applications.

  HACMP will not stop the managed resources; applications remain functional.

  HACMP does *not* manage the resources on these nodes.

  HACMP continues to run and RSCT remains functional.

Note: In HACMP 5.4, on a node that has Enhanced concurrent (ECM) volume groups, cluster services can be stopped with the resource groups placed in an unmanaged state. RSCT services will be left running so that ECM remains functional.

If you stop cluster services with this option, the resource groups that are active on this node go into unmanaged state. Once the resource group is in the unmanaged state, HACMP does not process any resource failures. This applies to hardware resources such as disks and adapters as well as any managed applications.

Refer to the section Procedure for Starting Cluster Services for information on reintegrating a node on which the cluster services were stopped back into the cluster.

Stop now, on system restart or both	Indicate whether you want the cluster services to stop **now**, at **restart** (when the operating system reboots), or on **both** occasions. If you select **restart** or **both**, the entry in the **/etc/inittab** file that starts cluster services is removed. Cluster services will no longer come up automatically after a reboot.
BROADCAST cluster shutdown?	Indicate whether you want to send a broadcast message to users before the cluster services stop. If you specify **true**, a message is broadcast on all cluster nodes.

5. Press Enter. The system stops the cluster services on the nodes specified.

If the stop operation fails, check the **/tmp/cspoc.log** file for error messages. This file contains the command execution status of the C-SPOC command executed on each cluster node.

Note: After stopping cluster services, you must wait a minimum of two minutes for the RSCT to quiesce before starting cluster services. If you are using HAGEO, wait for a minimum of four minutes.

Stopping HACMP Cluster Services without Stopping Applications

In HACMP 5.4, in addition to other ways to stop HACMP cluster services, you have a clear way of stopping cluster services *without stopping services and applications*.

Note: Prior to HACMP 5.4, stopping HACMP cluster services when it does *not* react to application failures was referred to as *forcing down* the cluster services. While “forcing down the cluster services” was desirable in many instances, in some cases, HACMP’s actions on resource groups after a force down left the applications unnecessarily inactive. In HACMP 5.4, this operation — stopping HACMP cluster services without disrupting the applications — is handled consistently with what you choose to do.

To stop cluster services without stopping your applications:

1. Enter the fastpath `smit cl_admin or smitty hacmp.`
2. Select **System Management (C-SPOC)** and press Enter.
3. Select **Manage HACMP Services > Stop Cluster Services** and press Enter.
4. Choose **Unmanage Resource Groups**.

No matter what type of resource group you have, if you stop cluster services on the node on which this group is active and do *not* stop the application that belongs to the resource group, HACMP puts the group into an UNMANAGED state and keeps the application running according to your request.

The resource group that contains the application remains in the UNMANAGED state (until you tell HACMP to start managing it again) and the application continues to run. While in this condition, HACMP and the RSCT services continue to run, providing services to ECM VGs that the application servers may be using.

You can tell HACMP to start managing it again either by restarting Cluster Services on the node, or by using SMIT to move the resource group to a node that is actively managing its resource groups.

If you have instances of replicated resource groups using the Extended Distance capabilities of the HACMP/XD product, the UNMANAGED SECONDARY state is used for resource groups that were previously in the ONLINE SECONDARY state.

You can view the new states of the resource groups using the cluster utilities **clstat** and **clRGinfo**.

You can dynamically reconfigure (DARE) the cluster configuration while some cluster nodes have resource groups in the unmanaged state.

Warning about Placing Resource Groups in an Unmanaged State

When you stop cluster services on a node and place resource groups in an UNMANAGED state, HACMP stops managing the resources on that node. HACMP will *not* react to the individual resource failures, application failures, or even if the node crashes.

Because the resources of a system are *not* highly available when you place resource groups in an unmanaged state, HACMP 5.4 prints a message periodically that the node has suspended managing the resources.

The ability to stop a node and place resource groups in an UNMANAGED state is intended for use during brief intervals for applying updates or for maintenance of the cluster hardware or software.

When You May Want to Stop HACMP Cluster Services without Stopping Applications

In general, HACMP cluster services are rarely the cause of problems in your configuration. However, you may still want to stop HACMP cluster services on one or more nodes, for example, while troubleshooting a problem or performing maintenance work on a node.

Also, you may want to stop HACMP cluster services from running without disrupting your application if you expect that your activities will interrupt or stop applications or services. During this period of time, you do *not* want HACMP to react to any planned application “failures” and cause a resource group to move to another node. Therefore, you may want to remove HACMP temporarily from the picture.

Abnormal Termination of Cluster Manager Daemon

The AIX source controller subsystem monitors the cluster manager daemon process. If the controller detects that the Cluster Manager daemon has exited abnormally (without being shut down using the **clstop** command), it executes the **/usr/es/sbin/cluster/utilities/clexit.rc** script to halt the system. This prevents unpredictable behavior from corrupting the data on the shared disks. See the **clexit.rc** man page for additional information.

The **clexit.rc** script creates an AIX 5L error log entry. Here is an example showing the long output:

```
LABEL:          OPMSG
IDENTIFIER:     AA8AB241

Date/Time:       Fri Jan  7 10:44:46
Sequence Number: 626
Machine Id:      000001331000
```

```
Node Id:         ppstest8
Class:           O
Type:            TEMP
Resource Name:   OPERATOR

Description
OPERATOR NOTIFICATION

User Causes
ERRLOGGER COMMAND

        Recommended Actions
        REVIEW DETAILED DATA

Detail Data
MESSAGE FROM ERRLOGGER COMMAND
clexit.rc : Unexpected termination of clstrmgrES
```

The **clexit.rc** error message in short form looks like this:

```
AA8AB241   0107104400 T O OPERATOR       OPERATOR NOTIFICATION
```

WARNING: Never use the **kill -9** command on the **clstrmgr** daemon. Using the **kill** command causes the **clstrmgr** daemon to exit abnormally. This causes the System Resource Controller (SRC) facility to run the script **/usr/es/sbin/cluster/utilities/clexit.rc**, which halts the system immediately and causes the surviving nodes to initiate fallover.

You can modify the file **/etc/cluster/hacmp.term** to change the default action after an abnormal exit. The **clexit.rc** script checks for the presence of this file, and if you have made it executable, the instructions there will be followed instead of the automatic halt called by **clexit.rc**. Please read the caveats contained in the **/etc/cluster/hacmp.term** file, however, before making any modifications.

AIX 5L Shutdown and Cluster Services

If you prefer to have resources taken over, then prior to issuing the AIX 5L **shutdown** command, stop HACMP cluster services with the **Move Resource Groups** option.

When the AIX operating system is shutdown on a node where the HACMP services are active, based on the command line flags that are passed to the shutdown command, the Cluster Manager either recovers the resource groups on a takeover node or simply leaves them in the offline state.

If you issue a shutdown command with "-F or -r" or a combination thereof, the resource groups are taken to the offline state. Resource groups will not fallover to the takeover nodes. The intent is that when the node starts backup, it may start the resource group on the same node.

If the shutdown command is issued with other options (such as -h), the node may not restart. In this case, HACMP will move the resource group to a takeover node.

Note: Using any other method of shutting down the AIX operating system (such as a halt command) or if the AIX operating system crashes results in HACMP recovering the failed application to a takeover node.

Stopping HACMP Cluster Services and RSCT

HACMP 5.4 manages the RSCT services automatically. When users stop cluster services using the Move Resource Group option, the RSCT services are stopped after all the resources and applications on the node are released. When users select the Unmanage Resource Group option to stop the cluster services, the Cluster Manager puts the resource groups into the UNMANAGED state but continues to run under the covers thus leaving the RSCT services up and running under this condition

One of the reasons that HACMP does *not* stop the RSCT services from running when you stop cluster services is because *not* only HACMP but also the Enhanced Concurrent Mode (ECM) volume groups use RSCT services. Stopping RSCT services would vary off the ECM volume group and would affect the application that is using it.

There could be rare cases when you need to stop RSCT, for example, to perform an RSCT upgrade. If you need to upgrade RSCT, you can stop and restart it, using SMIT options under the **HACMP Problem Determination Tools** menu. For the steps needed to stop, restart and upgrade RSCT, see the *Troubleshooting Guide.*

Maintaining Cluster Information Services

The cluster services on clients consist solely of the **clinfoES** daemon, which provides clients with status information about the cluster.

Note that the **/etc/inittab** file is modified when the HACMP software is installed to start the **clinfoES** daemon whenever the system is rebooted.

The Cluster Information Daemon (**clinfo**) retrieves information about the cluster configuration and the state of the cluster, topology and resources from the Management Information Base (MIB) and the Cluster Manager on local or remote nodes. The Cluster Manager updates the MIB with this information.

The **clinfo** daemon populates internal, dynamically allocated data structures with information for each cluster. The cluster(s) can be any combination of local or remote. The **clinfo** daemon calls the **clinfo.rc** script in response to cluster changes.

Starting Clinfo on a Client

Use the **/usr/es/sbin/cluster/etc/rc.cluster** script or the **startsrc** command to start **clinfo** on a client, as shown below:

```
/usr/es/sbin/cluster/etc/rc.cluster
```

You can also use the standard AIX 5L **startsrc** command:

```
startsrc -s clinfoES
```

Stopping Clinfo on a Client

Use the standard AIX 5L **stopsrc** command to stop **clinfo** on a client machine:

```
stopsrc -s clinfoES
```

Enabling Clinfo for Asynchronous Event Notification

In previous versions of HACMP, **clinfo** periodically polled the SNMP process for information. In HACMP 5.3 and up, **clinfo** only obtains data from SNMP when it is requested. You can optionally choose to have **clinfo** receive notification of events as asynchronous messages (traps).

Only one SNMP application can receive traps. If you are running NetView, you can*not* enable **clinfo** to receive traps.

To enable asynchronous event notification:

1. Start **clinfo** with the **-a** option, by entering the following:

   ```
   chssys -s clinfoES -a "-a".
   ```

2. Verify that the SRC has the correct command line arguments for **clinfo**, by entering the following:

   ```
   lssrc -Ss clinfoES | awk -F: '{print $3}'
   ```

3. Edit the **/etc/snmpd.conf** file on the nodes that will send traps. As installed, traps are directed to the loopback address. (**clinfo** receives those traps generated by the Cluster Manager on the same node). See the comments at the beginning of the **/etc/snmpd.conf** file for a description of all fields.

   Note: The default version of the **snmpd.conf** file for AIX 5L v.5.2 and AIX 5L v. 5.3 is **snmpdv3.conf**.

   See the AIX documentation for full information on the **snmpd.conf** file. Version 3 has some differences from Version 1.

   a. Find the trap line at the end of the file. It looks like this:

      ```
      view    1.17.2  system enterprises view
      trap   public  127.0.0.1       1.2.3   fe      # loopback
      ```

   b. Add trap lines as desired. Multiple **clinfo** processes can receive traps from the Cluster Manager. Make sure that the "`1.2.3 fe`" field is unique.

      An entry may look like the following example, with two more trap lines added:

      ```
      trap   public  127.0.0.1       1.2.3   fe      #loopback
      trap   public  123.456.789.1                   #adam
      trap   public  123.456.789.2                   #eve
      ```

   c. Stop and restart the **snmpd** process on the hosts where you made the changes in the **snmpd.conf** file:

      ```
      stopsrc -s snmpd
      startsrc -s snmpd
      ```

Gratuitous ARP Support

If you are using IPAT via IP Aliases, make sure all your clients support the gratuitous ARP functionality of TCP/IP. For more information, see Steps for Changing the Tuning Parameters of a Network Module to Custom Values in Chapter 13: Managing the Cluster Topology.

Chapter 10: Monitoring an HACMP Cluster

This chapter describes tools you can use to monitor an HACMP cluster.

You can use either ASCII SMIT or WebSMIT to configure and manage the cluster and view interactive cluster status. Starting with HACMP 5.4, you can also use WebSMIT to navigate, configure and view the status of the and graphical displays of the running cluster. For more information about WebSMIT, see Chapter 2: Administering a Cluster Using WebSMIT.

Note: The default locations of log files are used in this chapter. If you redirected any logs, check the appropriate location.

The main topics in this chapter include:

- Periodically Monitoring an HACMP Cluster
- Monitoring a Cluster with HAView
- Monitoring Clusters with Tivoli Distributed Monitoring
- Monitoring Clusters with clstat
- Monitoring Applications
- Monitoring Applications
- Displaying an Application-Centric Cluster View
- Using Resource Groups Information Commands
- Using HACMP Topology Information Commands
- Monitoring Cluster Services
- HACMP Log Files.

Periodically Monitoring an HACMP Cluster

By design, HACMP provides recovery for various failures that occur within a cluster. For example, HACMP can compensate for a network interface failure by swapping in a standby interface. As a result, it is possible that a component in the cluster has failed and that you are unaware of the fact. The danger here is that, while HACMP can survive one or possibly several failures, *each failure that escapes your notice threatens a cluster's ability to provide a highly available environment, as the redundancy of cluster components is diminished.*

To avoid this situation, you should customize your system by adding event notification to the scripts designated to handle the various cluster events. You can specify a command that sends you mail indicating that an event is about to happen (or that an event has just occurred), along with information about the success or failure of the event. The mail notification system enhances the standard event notification methods.

In addition, HACMP offers application monitoring capability that you can configure and customize in order to monitor the health of specific applications and processes.

Use the AIX 5L Error Notification facility to add an additional layer of high availability to an HACMP environment. You can add notification for failures of resources for which HACMP does *not* provide recovery by default. The combination of HACMP and the high availability features built into the AIX 5L system keeps single points of failure to a minimum; the Error Notification facility can further enhance the availability of your particular environment. See the chapter on Configuring AIX 5L for HACMP in the *Installation Guide* for suggestions on customizing error notification.

See Chapter 7: Planning for Cluster Events in the *Planning Guide* for detailed information on predefined events and on customizing event handling. Also, be sure to consult your worksheets, to document any changes you make to your system, and to periodically inspect the key cluster components to make sure they are in full working order.

Automatic Cluster Configuration Monitoring

Verification automatically runs on one user-selectable HACMP cluster node once every 24 hours. By default, the first node in alphabetical order runs the verification at midnight. If verification finds errors, it warns about recent configuration issues that might cause problems at some point in the future. HACMP stores the results of the automatic monitoring on every available cluster node in the **/var/hacmp/log/clutils.log** file.

If cluster verification detects some configuration errors, you are notified about the potential problems:

- The exit status of verification is published across the cluster along with the information about cluster verification process completion.
- Broadcast messages are sent across the cluster and displayed on **stdout**. These messages inform you about detected configuration errors.
- A **cluster_notify** event runs on the cluster and is logged in **hacmp.out** (if cluster services is running).

More detailed information is available on the node that completes cluster verification in **/var/hacmp/clverify/clverify.log** file. If a failure occurs during processing, error messages and warnings clearly indicate the node and reasons for the verification failure.

Tools for Monitoring an HACMP Cluster

HACMP supplies tools for monitoring a cluster. These are described in subsequent sections:

- The **HAView** utility extends Tivoli NetView services so you can monitor HACMP clusters and cluster components from a single node. Using HAView, you can also view the full cluster event history in the **/usr/es/sbin/cluster/history/cluster.*mmddyyyy*** file. The event history (and other cluster status and configuration information) is accessible through Tivoli NetView's menu bar. For more information, see Monitoring a Cluster with HAView.
- **Cluster Monitoring with Tivoli** allows you to monitor clusters and cluster components and perform cluster administration tasks through your Tivoli Framework console. For more information, see Monitoring Clusters with Tivoli Distributed Monitoring.
- **clstat** (the **/usr/es/sbin/cluster/clstat** utility) reports the status of key cluster components—the cluster itself, the nodes in the cluster, the network interfaces connected to the nodes, the service labels, and the resource groups on each node.

- WebSMIT displays cluster information using a slightly different layout and organization. Cluster components are displayed along their status. Expanding the item reveals additional information about it, including the network, interfaces and active resource groups.

For more information, see Monitoring Clusters with clstat.

- **Application Monitoring** allows you to monitor specific applications and processes and define action to take upon detection of process death or other application failures. Application monitors can watch for the successful startup of the application, check that the application runs successfully after the stabilization interval has passed, or monitor both the startup and the long-running process. For more information, see Monitoring Applications.
- SMIT and WebSMIT give you information on the cluster.

  You have the ability to see the cluster from an application-centric point of view.

  - The **HACMP Resource Group and Application Management** panel in SMIT has an option to **Show Current Resource Group and Application State**. The SMIT panel **Show All Resources by Node or Resource Group** has an option linking you to the **Show Current Resource Group and Application State** panel.
  - Using the WebSMIT version lets you expand and collapse areas of the information. Colors reflect the state of individual items (for example, green indicates online).

  For more information, see Displaying an Application-Centric Cluster View.

  The **System Management (C-SPOC) >Manage HACMP Services > Show Cluster Services** SMIT panel shows the status of the HACMP daemons.
- The **Application Availability Analysis** tool measures uptime statistics for applications with application servers defined to HACMP. For more information, see Measuring Application Availability.
- The **clRGinfo** and **cltopinfo** commands display useful information on resource group configuration and status and topology configuration, respectively. For more information, see Using Resource Groups Information Commands.
- **Log files** allow you to track cluster events and history: The **/usr/es/adm/cluster.log** file tracks cluster events; the **/tmp/hacmp.out** file records the output generated by configuration scripts as they execute; the **/usr/es/sbin/cluster/history/cluster.*mmddyyyy*** log file logs the daily cluster history; the **/tmp/cspoc.log** file logs the status of C-SPOC commands executed on cluster nodes. You should also check the RSCT log files. For more information, see HACMP Log Files.

In addition to these cluster monitoring tools, you can use the following:

- The **Event Emulator** provides an emulation of cluster events. For more information, see the section on Emulating Events in the *Concepts Guide*.
- The **Custom Remote Notification** utility allows you to define a notification method through the SMIT interface to issue a customized page in response to a cluster event. In HACMP 5.3 and up, you can also send text messaging notification to any address including a cell phone. For information and instructions on setting up pager notification, see the section on Configuring a Custom Remote Notification Method in the *Planning Guide*.

Monitoring a Cluster with HAView

HAView is a cluster monitoring utility that allows you to monitor HACMP clusters using NetView for UNIX. Using Tivoli NetView, you can monitor clusters and cluster components across a network from a single management station.

HAView creates and registers Tivoli NetView objects that represent clusters and cluster components. It also creates submaps that present information about the state of all nodes, networks, network interfaces, and resource groups associated with a particular cluster. This cluster status and configuration information is accessible through Tivoli NetView's menu bar.

HAView monitors cluster status using the Simple Network Management Protocol (SNMP). It combines periodic polling and event notification through traps to retrieve cluster topology and state changes from the HACMP management agent, the Cluster Manager.

You can view cluster event history using the HACMP Event Browser and node event history using the Cluster Event Log. Both browsers can be accessed from the Tivoli NetView menu bar. The **/usr/es/sbin/cluster/history/cluster.mmddyyyy** file contains more specific event history. This information is helpful for diagnosing and troubleshooting fallover situations. For more information about this log file, see Chapter 2: Using Cluster Log Files in the *Troubleshooting Guide*.

HAView Installation Requirements

HAView has a client/server architecture. You must install both an HAView server image and an HAView client image, on the same machine or on separate server and client machines. For information about installation requirements, see *Installation Guide.*

HAView File Modification Considerations

Certain files need to be modified in order for HAView to monitor your cluster properly. When configuring HAView, you should check and edit the following files:

- **haview_start**
- **clhost**
- **snmpd.conf** or **snmpdv3.conf**

haview_start File

You must edit the **haview_start** file so that it includes the name of the node that has the HAView server executable installed. This is how the HAView client knows where the HAView server is located. Regardless of whether the HAView server and client are on the same node or different nodes, you are required to specify the HAView server node in the **haview_start** file.

The **haview_start** file is loaded when the HAView client is installed and is stored in **/usr/haview**. Initially, the **haview_start** file contains only the following line:

```
"${HAVIEW_CLIENT:-/usr/haview/haview_client}" $SERVER
```

You must add the following line to the file:

```
SERVER="${SERVER:-<your server name>}"
```

For example, if the HAView server is installed on *mynode*, the edited **haview_start** file appears as follows:

```
SERVER="${SERVER:-mynode}"
"${HAVIEW_CLIENT:-/usr/haview/haview_client}" $SERVER
```

where *mynode* is the node that contains the HAView server executable.

Note: If you have configured a persistent node IP label on a node on a network in your cluster, it maintains a persistent "node address" on the node – this address can be used in the **haview_start** file.

clhosts File

HAView monitors a cluster's state within a network topology based on cluster-specific information in the **/usr/es/sbin/cluster/etc/clhosts** file. The **clhosts** file must be present on the Tivoli NetView management node. Make sure this file contains the IP address or IP label of the service and/or base interfaces of the nodes in each cluster that HAView is to monitor.

Make sure that the hostname and the service label of your Tivoli NetView nodes are exactly the same. (If they are *not* the same, add an alias in the **/etc/hosts** file to resolve the name difference.)

WARNING: If an invalid IP address exists in the **clhosts** file, HAView will fail to monitor the cluster. Make sure the IP addresses are valid, and there are no extraneous characters in the **clhosts** file.

snmpd.conf File

The Tivoli NetView management node must also be configured in the list of trap destinations in the **snmpd.conf** files on the cluster nodes of all clusters you want it to monitor. This makes it possible for HAView to utilize traps in order to reflect cluster state changes in the submap in a timely manner. Also, HAView can discover clusters *not* specified in the **clhosts** file on the nodes in another cluster.

Note: The default version of the **snmpd.conf** file for AIX 5L v.5.2 and AIX 5L v. 5.3 is **snmpdv3.conf**.

The format for configuring trap destinations is as follows:

```
trap <community name> <IP address of Tivoli NetView management node> 1.2.3 fe
```

For example, enter:

```
trap  public  140.186.131.121 1.2.3 fe
```

Note the following:

- You can specify the name of the management node instead of the IP address.
- You can include multiple trap lines in the **snmpd.conf** file.

Note: HACMP now supports a SNMP Community Name other than "public." If the default SNMP Community Name has been changed in **/etc/snmpd.conf** to something different from the default of "public" HACMP will function correctly. The SNMP Community Name used by HACMP will be the first name found that is *not* "private" or "system" using the **lssrc -ls snmpd** command.

Clinfo will also get the SNMP Community Name in the same manner. Clinfo will still support the **-c** option for specifying SNMP Community Name but its use is *not* required. The use of the **-c** option is considered a security risk because doing a **ps** command could find the SNMP Community Name. If it is important to keep the SNMP Community Name protected, change permissions on **/tmp/hacmp.out**, **/etc/snmpd.conf**, **/smit.log** and **/usr/tmp/snmpd.log** to *not* be world-readable.

See the AIX documentation for full information on the **snmpd.conf** file. Version 3 has some differences from Version 1.

Tivoli NetView Hostname Requirements for HAView

The following hostname requirements apply to using HAView in a Tivoli NetView environment. If you change the hostname of a network interface, the Tivoli NetView database daemons and the default map are affected.

Hostname Effect on the Tivoli NetView Daemon

The hostname required to start Tivoli NetView daemons must be associated with a valid interface name or else Tivoli NetView fails to start.

Hostname Effect on the Tivoli NetView Default Map

If you change the hostname of the Tivoli NetView client, the new hostname does *not* match the original hostname referenced in the Tivoli NetView default map database and Tivoli NetView will *not* open the default map. Using the Tivoli NetView **mapadmin** command, you need to update the default map (or an invalid map) to match the new hostname.

See the *Tivoli NetView Administrator's Guide* for more information about updating or deleting an invalid Tivoli NetView map.

Starting HAView

Once you have installed the HAView client and server, HAView is started and stopped when you start or stop Tivoli NetView. However, before starting Tivoli NetView/HAView, check the management node as follows:

- Make sure both client and server components of HAView are installed. See the installation or migration chapters in the *Installation Guide* for more information.
- Make sure access control has been granted to remote nodes by running the **xhost** command with the plus sign (+) or with specified nodes:

  `xhost +` (to grant access to all computers)

  or, to grant access to specific nodes only:

  `xhost <computers to be given access>`
- Make sure the DISPLAY variable has been set to the monitoring node and to a label that can be resolved by and contacted from remote nodes:

  `export DISPLAY=<monitoring node>:0.0`

These actions allow you to access HACMP SMIT panels using the HAView Cluster Administration option.

After ensuring these conditions are set, type the following to start Tivoli NetView:

```
/usr/OV/bin/nv6000
```

Refer to the *Tivoli NetView User's Guide for Beginners* for further instructions about starting Tivoli NetView.

When Tivoli NetView starts, HAView creates objects and symbols to represent a cluster and its components. Through submaps, you can view detailed information about these components.

HAView places the Clusters symbol (shown below) on the Tivoli NetView map after Tivoli NetView starts. The Clusters symbol is added to the Netview Root map and is placed alongside the Tivoli NetView Collections symbol and other symbols:

HAView Clusters Symbol

Viewing Clusters and Components

To see which clusters HAView is currently monitoring, double-click the Clusters symbol. The Clusters submap appears. You may see one or more symbols that represent specific clusters. Each symbol is identified by a label indicating the cluster's name. Double-click a cluster symbol to display symbols for nodes, networks, and resource groups within that cluster.

Note that the cluster status symbol may remain unknown until the next polling cycle, even though the status of its cluster components is known. See Customizing HAView Polling Intervals for more information about the default intervals and how to change them using SMIT.

You can view component details at any time using the shortcut **ctrl-o**. See Obtaining Component Details in HAView for information and instructions.

Read-Write and Read-Only NetView Maps

Normally, you have one master monitoring station for Tivoli NetView/HAView. This station is supplied with new information as cluster events occur, and its map is updated so it always reflects the current cluster status.

In normal cluster monitoring operations, you will probably *not* need to open multiple Tivoli NetView stations on the same node. If you do, and you want the additional stations to be updated with current cluster status information, you must be sure they use separate maps with different map names. For more information on multiple maps and changing map permissions, see the *Tivoli NetView Administrator's Guide*.

Interpreting Cluster Topology States

When using HAView to view cluster topology, symbols for clusters and cluster components such as nodes and networks are displayed in various colors depending on the object's state. The following table summarizes colors you may see when monitoring a cluster. (For information about the resource group symbol colors, see Interpreting Resource Group Symbol Colors in HAView.

Status	Meaning	Symbol Color	Connection Color (network submap)
Critical	The object has failed or is *not* functioning. If the symbol is a node or network, the node or network is DOWN.	**Red**	**Red**
Normal	The object is functioning correctly. If the symbol is a node object, the node is UP.	**Green**	**Black**
Marginal	Some object functions are working correctly; others are *not*.	**Yellow**	
Unknown	The object's state can*not* be determined. It may *not* be currently monitored by HAView.	**Blue**	**Blue**

You can select **Legend** at any time from the Help pull-down menu to view Tivoli NetView and HAView symbols and to understand their associative colors.

The Navigation Tree and Submap Windows

In addition to the submap window, the Tivoli NetView Navigation Tree Window can help you keep track of your current location in the HAView hierarchy. Press the Tree button to see the Navigation Tree Window. In the Navigation Tree, the blue outline indicates where you are in the map, that is, which submap you are in.

The Symbols Legend

At any time, you can select **Legend** from the Help pull-down menu to view all Tivoli NetView and HAView symbols and the meanings of the symbols and their various colors.

The Help Menu

To view help topics, select **Help > Index > Tasks > HAView Topics**.

Viewing Networks

To view the state of the nodes and addresses connected to a network associated with a specific cluster, double-click a network symbol in the specific Cluster submap. A network submap appears displaying symbols for all nodes connected to the network. The symbols appear in a color that indicates the nodes' current state. The vertical line representing a network is called the network connection. Its color indicates the status of the connection between the node and the network.

See Interpreting Cluster Topology States in the next section for a table of symbol colors and how they reflect a cluster and its components' state.

Viewing Nodes

To view the state of nodes associated with a particular network, double-click a network symbol. A submap appears displaying all nodes connected to the network. Each symbol's color indicates the associated node's current state.

You can also view the state of any individual node associated with a cluster by double-clicking on that node's symbol in the specific cluster submap.

Viewing Addresses

To view the status of addresses serviced by a particular node, double-click a node symbol from either a cluster or network submap. A submap appears displaying symbols for all addresses configured on a node. Each symbol's color indicates the associated address's current state.

When you view interfaces in a node submap from a network submap, all interfaces relating to that node are shown, even if they are *not* related to a particular network.

Viewing Resource Groups and Resources

Note: In HACMP 5.2 and up, resource groups are displayed in HAView as type "unknown."

Resource Group Ownership Symbol

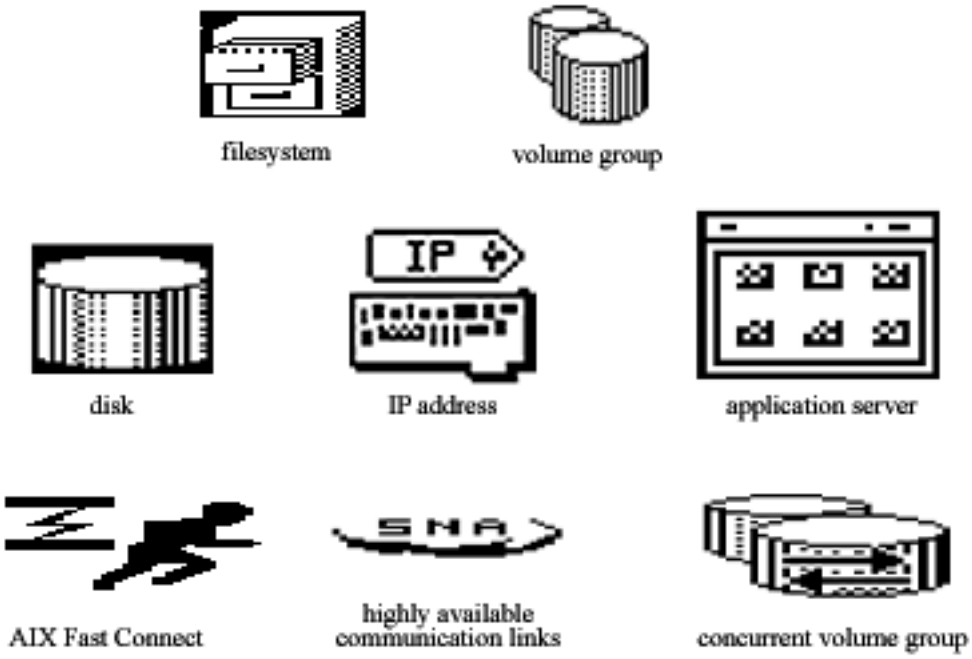
HAView indicates the current ownership of a resource group in both the Resource Group and Node submaps by showing the owner node together with the resource group as follows.

In the *Resource Group* submap, ownership is shown with this symbol:

Resource Group Ownership Symbol in Resource Group Submap

Resource Submap—Individual Resource Symbols

The HAView Resource Group submap displays all the individual resources configured as part of a given resource group. Each type of resource has its own symbol, as shown in the following figure:



Symbols for Individual Resource Types

Remember that individual resource symbols always appear in the blue (unknown) state, regardless of their actual state; HAView does *not* monitor the status of individual resources, only their presence and location.

Interpreting Resource Group Symbol Colors in HAView

Each symbol's color indicates the current state of the associated resource group, as follows:

Resource Group Status	Symbol Color	What Is Occurring
Online/UP	**green**	The resource group is currently operating properly on one or more nodes in the cluster
Offline/DOWN	**red**	The resource group is *not* operating in the cluster and is *not* in an error condition.
Acquiring	**yellow**	The resource group is currently trying to come up on one of the nodes in the cluster.

Releasing	**yellow**	The resource group is in the process of being released from the ownership of a node.
Error	**blue**	The resource group has reported an error condition, and intervention is required.
Unknown	**blue**	The resource group's current status can*not* be obtained, possibly due to a loss of communication between the monitoring node and the cluster.

Obtaining Component Details in HAView

Tivoli NetView dialog boxes allow you to view detailed information about a cluster object. A dialog box can contain information about a cluster, network, node, network interface, or resource group, or about cluster events. You can access an object's dialog box using the Tivoli NetView menu bar or the Object Context menu, or by pressing **ctrl-o** at any time.

To view details about a cluster object using the Tivoli NetView menu bar:

1. Click on an object in any submap.
2. Select the **Modify/Describe** option from the Tivoli NetView Edit menu.
3. Select the **Object** option. An Object Description dialog window appears.
4. Select **HAView for AIX** and click on **View/Modify Object Attributes**. An Attributes dialog window appears.

   You can view dialog boxes for more than one object simultaneously by either clicking the left mouse button and dragging to select multiple objects, or by pressing the **Alt** key and clicking on all object symbols for which you want more information.

To view details about a cluster object using the Object Context menu:

1. Click on an object in any submap.
2. Click on the symbol you have highlighted to display the object context menu, using **Button 3** on a three-button mouse **Button 2** on a two-button mouse.
3. Select **Edit** from the object context menu.
4. Select **Modify/Describe** from the Edit cascade menu.
5. Select the **Object** option. An Object Description dialog window appears.
6. Select **HAView for AIX** and click on **View/Modify Object Attributes**. An Attributes dialog window appears.

Customizing HAView Polling Intervals

To ensure that HAView is optimized for system performance and reporting requirements, you can customize these two parameters:

- The polling interval (in seconds) at which HAView polls the HACMP clusters to determine if cluster configuration or object status has changed. The default is 60 seconds.

- The polling interval (in minutes) at which HAView polls the **clhosts** file to determine if new clusters have been added. The default for Cluster Discovery polling is 120 minutes.

You can change the HAView polling intervals using the SMIT interface as follows:

1. On the HAView server node, open a SMIT panel by typing: `smitty haview`. The **Change/Show Server Configuration** window opens.
2. Enter the polling interval numbers you want (between 1 and 32000) and press OK.

Note: If the **snmpd.conf** file is *not* properly configured to include the Tivoli NetView server as a trap destination, HAView can detect a trap that occurs as a result of a cluster event, but information about the network topology may *not* be timely. Refer back to the section HAView File Modification Considerations for more information on the **snmpd.conf** file.

Removing a Cluster from HAView

If a cluster does *not* respond to status polling, you can use the **Remove Cluster** option to remove the cluster from the database. To remove a cluster, the cluster state must be UNKNOWN, as represented by a blue cluster symbol. If the cluster is in any other state, the **Remove Cluster** option is disabled.

WARNING: The **Remove Cluster** option is the only supported way to delete HAView objects from submaps. Do *not* delete an HAView symbol (cluster or otherwise) through the Delete Object or Delete Symbol menu items. If you use these menu items, HAView continues to poll the cluster.

When you remove a cluster, the following actions occur:

- The cluster name is removed from the Tivoli NetView object database and HAView stops polling the cluster.
- The symbol for the cluster is deleted.
- The symbols for all child nodes, networks, addresses, and resource groups specific to that cluster are deleted.

If you are removing the cluster permanently, remember to remove the cluster addresses from the **/usr/es/sbin/cluster/etc/clhosts** file. If you do *not* remove the cluster addresses from the **clhosts** file, new cluster discovery polling continues to search for the cluster.

To remove a cluster:

1. Click on the cluster symbol you wish to remove. The cluster must be in an UNKNOWN state, represented by a blue cluster symbol.
2. Select **HAView** from the Tools pull-down menu.
3. Select **Remove Cluster** from the HAView cascade menu.

Using the HAView Cluster Administration Utility

HAView allows you to start a SMIT HACMP session to perform cluster administration functions from within the Tivoli NetView session. The administration session is run on an aixterm opened on the chosen node through a remote shell. You can open multiple sessions of SMIT HACMP while in HAView. You must have root permissions, or enter the root password, to open a SMIT panel.

Note: You can start an administration session for any node that is in an UP state (the node symbol is green). If you attempt to start an administration session when the state of the node is DOWN or UNKNOWN, no action occurs.

When bringing a node up, the HAView node symbol may show green before all resources are acquired. If you select the node symbol and attempt to open an administration session before all resources are acquired, you may receive an error.

Opening and Closing a Cluster Administration Session

To open a cluster administration session:

1. Click on an available node symbol (the one that is green).
2. Select the **Tools > HAView > Cluster Administration**.

   If you are a non-root user, an AIX 5L window appears prompting you to enter the root password. When the password is verified, a SMIT window opens.
3. Proceed with your tasks in SMIT.
4. Exit the Cluster Administration session; the aixterm session will also close.

Cluster Administration Notes and Requirements

Keep in mind the following considerations when using the Cluster Administration option:

- Be sure you have run the **xhost** command prior to starting Tivoli NetView, so that a remote node can start an aixterm session on your machine.
- Be sure you have set the DISPLAY variable to a label that can be resolved and contacted from remote nodes.
- For the cluster administration session to proceed properly, the current Tivoli NetView user (the account that started Tivoli NetView) must have sufficient permission and be authenticated to perform an **rsh** to the remote node in the **~/.rhosts** file or through Kerberos.
- If an IP Address Takeover (IPAT) occurs while a cluster administration session is running, the route between the remote node and the HAView monitoring node may be lost.

HAView Browsers

HAView provides two browsers that allow you to view the event history of a cluster, the Cluster Event Log and the HACMP Event Browser.

Cluster Event Log

Using the Cluster Event Log you can view the event history for a cluster as recorded by a specific node. The Log browser is accessible through the Tivoli NetView Tools menu, and is only selectable if an active node symbol is highlighted.

For more detailed information on a node's event history, log onto the specific node and check the cluster message log files. For more information on these logs see the Cluster Message Log Files section in Chapter 2: Using Cluster Log Files in the *Troubleshooting Guide*.

Note: To ensure that the header for the Cluster Event Log displays properly, install all of the Tivoli NetView fonts on your system.

To review a cluster event log:

1. Click on the node symbol for which you wish to view a Cluster Event Log.
2. Select **HAView** from the Tivoli NetView Tools menu.
3. Select the **Cluster Event Log** option.
4. Set the **number of events to view** field. You can use the up and down arrows to change this number or you can enter a number directly into the field. The possible range of values is 1 to 1000 records. The default value is 100.
5. Press the **Issue** button to generate the list of events. The message area at the bottom of the dialog box indicates when the list is done generating.

   When the list is done generating, the dialog box displays the following view-only fields:

Event ID	This field displays a numeric identification for each event that occurred on the cluster.
Node Name	The name of the node on which the event occurred.
Time	The date and time the event occurred. This field is in the format MM DD hh:mm:ss.
Description	A description of the event.

6. When you are finished, press the **Dismiss** button to close the dialog box.

HACMP Event Browser

HAView provides a Tivoli NetView browser that allows you to view the accumulative event history of a cluster. The browser shows the history of all nodes in the cluster, broadcast through an assigned primary node. If the primary node fails, another node assumes the primary role and continues broadcasting the event history.

The HACMP Event Browser provides information on cluster state events. A filter is used to block all redundant traps.

To view the HACMP Event Browser:

1. Select **HAView** from the Tivoli NetView Tools menu. The menu item is always active, and when selected will start a Tivoli NetView browser showing the event history for all active clusters.

2. Select the **HACMP Event Browser** option.The HACMP Event Browser appears. Note that only one instantiation of the Event Browser can be accessed at a time. See the *Tivoli NetView User's Guide for Beginners* for more information on the Tivoli NetView browser functions.
3. Select the **Close** option from the File menu of the HACMP Event Browser menu bar to close the browser.

When you exit the Event Browser, the HAView application restarts. At this time, the HACMP cluster icon turns blue, disappears, and then reappears.

Monitoring Clusters with Tivoli Distributed Monitoring

You can monitor the state of an HACMP cluster and its components and perform some cluster administration tasks through your Tivoli Framework enterprise management system.

In order to integrate HACMP with Tivoli, you must configure your HACMP cluster nodes as subscriber (client) nodes to the Tivoli server node, or Tivoli Management Region (TMR). Each cluster node can then maintain detailed node information in its local Tivoli database, which the TMR accesses for updated node information to display.

The following sections discuss how to monitor your cluster once you have set up the cluster nodes and Tivoli so that Tivoli can monitor the cluster. If you have *not* done this setup yet, see the Appendix on configuring Tivoli for HACMP in the *Installation Guide* for instructions.

Cluster Monitoring and Cluster Administration Options

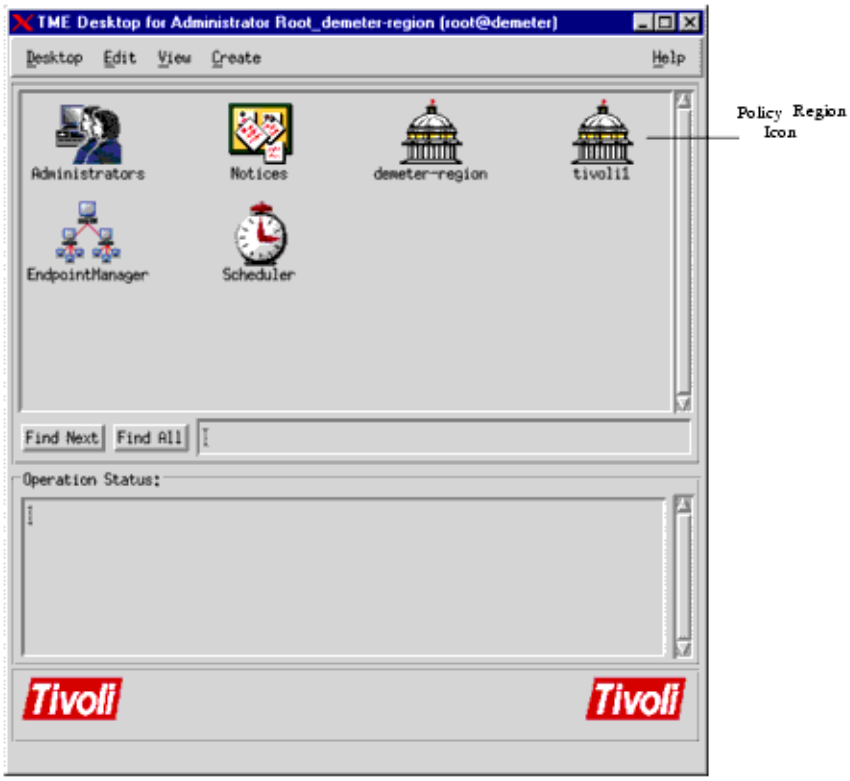
Using various windows of the Tivoli interface, you can *monitor* the following aspects of your cluster:

- Cluster state and substate
- Configured networks and network state
- Participating nodes and node state
- Configured resource group location and state
- Individual resource location (*not* state).

In addition, you can perform the following *cluster administration* tasks from within Tivoli:

- Start cluster services on specified nodes
- Stop cluster services on specified nodes
- Bring a resource group online
- Bring a resource group offline
- Move a resource group to another node.

The initial Tivoli NetView Desktop view is shown here:



Tivoli Desktop Initial Panel

Tivoli's thermometer icons provide a visual indication of whether components are up, down, in transition, or in an unknown or error state. From the window for a selected Policy Region, you can go a cluster's Indicator Collection window, which displays thermometer icons indicating the state of all cluster components.

The cluster status information shown by the thermometers is updated every three minutes by default or at another interval you specify. (Further information on changing the default polling interval appears later in this chapter. See Customizing HAView Polling Intervals.)

Note: The following sections provide information on monitoring an HACMP cluster through the Tivoli interface. Descriptions of Tivoli components and processes are provided here as needed, but for full information on installing, configuring, and using the Tivoli software itself, consult your Tivoli product documentation.

For complete details on setting up HACMP cluster monitoring with Tivoli, see the *Installation Guide*.

Using Tivoli to Monitor the Cluster

Once you have properly installed your **hativoli** files and defined your nodes to Tivoli, you can view information on the status of your HACMP cluster components.

When you monitor your cluster through Tivoli, you can access cluster information in both icon and text form, in a number of different Tivoli windows. The next few sections are meant to orient you to the flow of Tivoli cluster monitoring information.

Note: When HATivoli is unable to contact nodes in the cluster, or when all nodes are down, node status may *not* be displayed accurately. You should be aware that in the event that your last remaining cluster node goes down, Tivoli may still indicate that the cluster is up. This can occur when HACMP is unable to contact the Management Information Base (MIB) for updated information. In this case, the Tivoli display will show information *as of the last successful poll*.

Starting Tivoli

If Tivoli is *not* already running, start Tivoli by performing these steps on the TMR node:

1. Make sure access control has been granted to remote nodes by running the **xhost** command with the plus sign (+) or with specified nodes. This will allow you to open a SMIT window from Tivoli.

   If you want to grant access to all computers in the network, type:

   ```
   xhost +
   ```

   or, if you want to grant access to specific nodes only:

   ```
   xhost <computers to be given access>
   ```

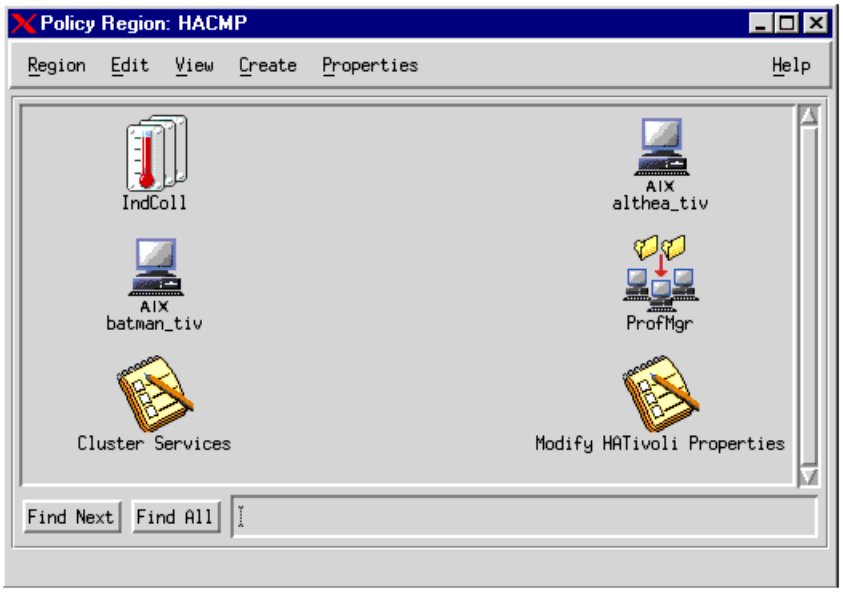
2. Also to ensure later viewing of SMIT windows, set DISPLAY=<*TMR node*>.
3. Run the command **. /etc/Tivoli/setup_env.sh** if it was *not* run earlier.
4. Enter **tivoli** to start the application.The Tivoli graphical user interface appears, showing the initial Tivoli Desktop window.

Note that there may be a delay as Tivoli adds the indicators for the cluster.

Tivoli Policy Regions

A Tivoli *Policy Region* groups together all entities related to a specific area of Tivoli monitoring. In this case, that area is the HACMP cluster or clusters. The HACMP Policy Region encompasses the nodes, clusters, indicator icons, and tasks related to your HACMP configuration.

Policy Region icons appear in the initial Tivoli Desktop window (shown in the preceding figure). Clicking on a Policy Region icon opens the Policy Region window, in which you see the thermometer icons of the Indicator Collections, as well as icons for the profiles and task libraries associated with the HACMP Policy Region.



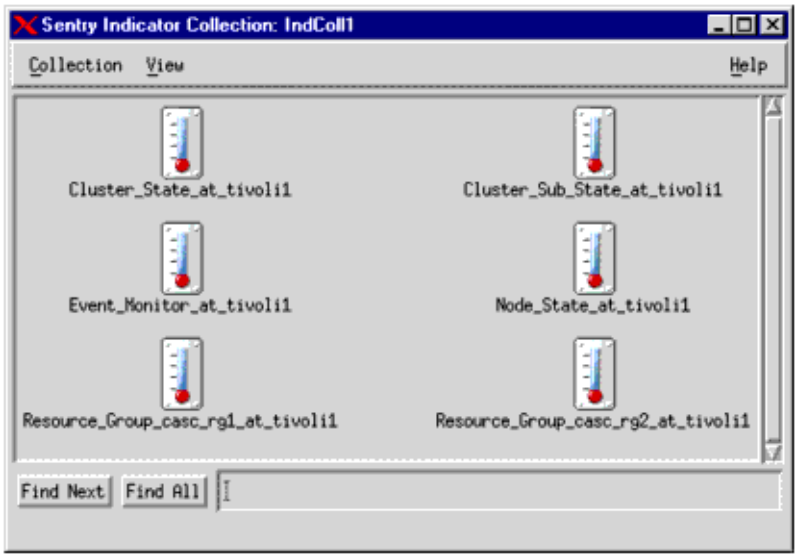
HACMP Policy Region Window

Tivoli Distributed Monitors and Indicator Collections

For each cluster, a group of Tivoli *distributed monitors* is created; these monitors query the HACMP cluster node at set intervals for information about the various cluster components. The group of monitors is associated with an *indicator collection* that displays the state of the cluster components. If a change is detected in the state of the cluster or one of its components, the distributed monitor takes action, changing an icon in the associated Indicator Collection window. This provides the Tivoli administrator with a visual representation of any changes in the status of the monitored items.

New monitors are added whenever new cluster components are configured. When cluster components are removed from the cluster configuration, the associated monitors are also removed.

The icon for cluster and cluster component status is a thermometer figure with varying levels of red color depending on the severity of the component status. When you click on a cluster's Indicator Collection icon in the Policy Region window, the Indicator Collection window appears showing status icons for that cluster's components.



Indicator Collection Window

Interpreting Indicator Displays for Various Cluster Components

The Indicator icons reflect varying degrees of severity of problems, depending on the height of the red color in the thermometer and the color-coded marker alongside it. The following tables list the indicator displays for various cluster component states:

CLUSTER STATE Indicator Display	Cluster State
Normal	UP
Fatal	DOWN
Severe	UNKNOWN

CLUSTER SUBSTATE Indicator Display	Cluster Substate
Normal	STABLE
Warning	UNSTABLE
Severe	RECONFIG
Critical	ERROR

NODE Indicator Display*	Node State(s)
Normal	All nodes ONLINE
Warning	One or more nodes OFFLINE

Note that node state is displayed in the Distributed Monitor Indicator Collection as a composite view of all nodes rather than an individual node view.

RESOURCE GROUP Indicator Display	Resource Group State
Normal	ONLINE
Warning	ACQUIRING
Warning	RELEASING
Critical	ERROR
Fatal	OFFLINE

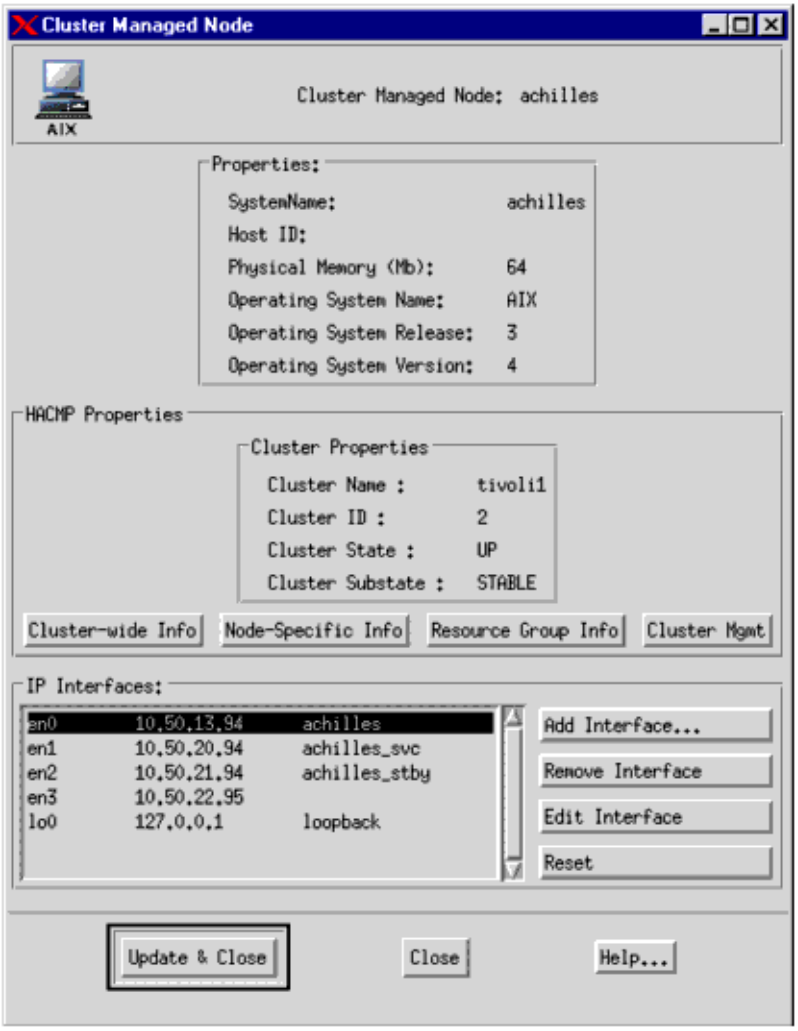
Viewing Cluster Information

The Cluster Managed Node window gives you all information about the current cluster topology and configuration.

The Properties section displays standard system properties information for the managed node, and the IP Interfaces section at the bottom shows standard IP Interface information for the node.

With the addition of the HACMP cluster monitoring feature, the standard Tivoli node icon is extended to capture and display additional information specific to your HACMP cluster. The items that you see in the HACMP Properties portion of the window are detailed in the following sections.

To view the Managed Node window, right-click on a node icon in the Policy Region window.



Cluster Managed Node Window

In the HACMP Properties portion of the Managed Node window shown above, you see the following four items of HACMP-specific top-level cluster information: Cluster name, Cluster ID, Cluster state, and Cluster substate.

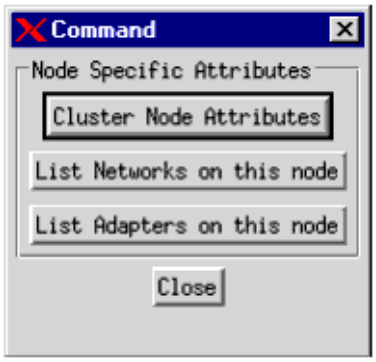
In addition, HACMP Properties buttons lead you to further cluster and component details. Selecting a button brings up a new popup window with options for retrieving specific information. These buttons and the choices within them are detailed below.

Cluster-Wide Information Button

Clicking the cluster-wide button gives you the Cluster Information window, shown below, with options to view details on the configuration and status information for the cluster as a whole.

Node-Specific Information Button

Clicking the node-specific button brings up the Node Specific Attributes window, from which you can access further information about the attributes, networks, and interfaces associated with the node.



Resource Group Information Button

Clicking the resource group information button brings up the Resource Group Information window; from there, you can access further details about the resource groups in your cluster and their associated resources and nodes.

Note: All resource groups are displayed in Tivoli as type “unknown.”

Cluster Management Button

Clicking the Cluster Management button brings up the Cluster Management window. From here, you can open a SMIT window to perform all of your normal cluster management tasks.

If you are a non-root user, you will be prompted to enter the root password after clicking the Open SMIT Window button. When the password is verified, a SMIT window opens and you can proceed with cluster management activity.

Besides having root permissions or entering the root password, in order to open a SMIT window from within Tivoli, you must have run the **xhost** command to grant access to remote nodes. See instructions in the section Starting Tivoli.

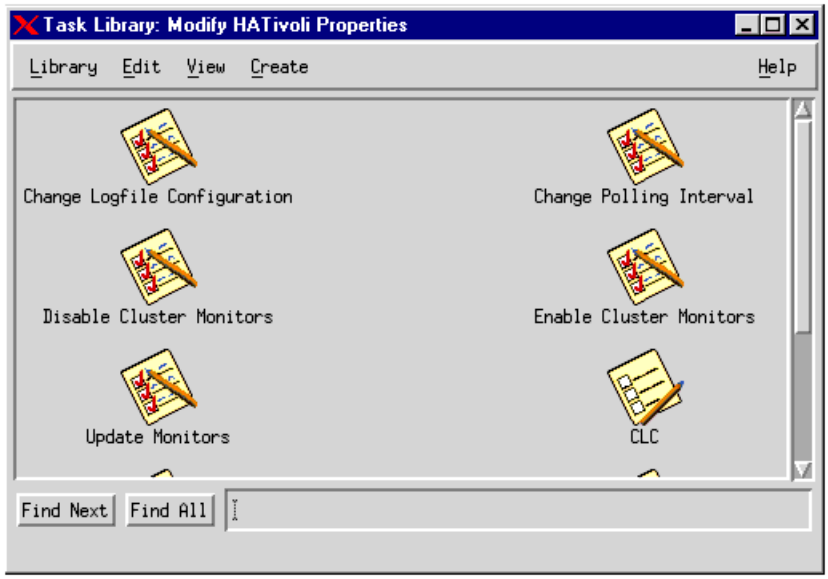
Customizing Polling Intervals

The Distributed Monitors poll the cluster nodes periodically for cluster topology and status changes. The default polling interval is three minutes. If this interval is too short for your particular cluster monitoring needs, you can change this interval through an HACMP Task found in the Modify HATivoli Properties Task Library. It is *not* recommended to make the polling interval shorter than the default.

As mentioned earlier, be aware that if your last remaining cluster node goes down, Tivoli may still indicate that the cluster is up. This can occur when HACMP is unable to contact the MIB for updated information. In this case, the Tivoli display will show information *as of the last successful poll*.

Modifying HATivoli Properties

The Modify HATivoli Properties Task Library window shown below contains options to perform cluster management tasks such as configuring, modifying, and deleting various items associated with the cluster, and refreshing the cluster view.



Modify HATivoli Properties Task Library

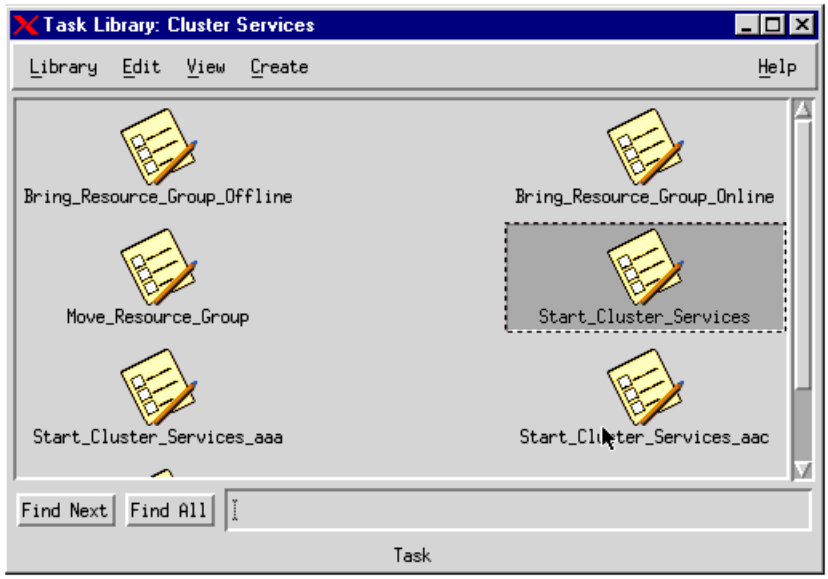
Click on the appropriate library to perform the desired tasks.

Using Tivoli to Perform Cluster Administration Tasks

You can perform several cluster administration tasks from within Tivoli: You can start or stop cluster services on a specified node, bring a resource group online or offline, and move a resource group from a specified node to a target node.

Note: In order to perform cluster administration tasks, you must have *admin* level privileges under Resource Roles in the HACMP policy region.

The cluster administration tasks are found in the Cluster Services Task Library, shown here.



Cluster Services Task Library

Overview of Steps for Configuring Cluster Administration Tasks

All cluster administration tasks performed through Tivoli require these basic steps:

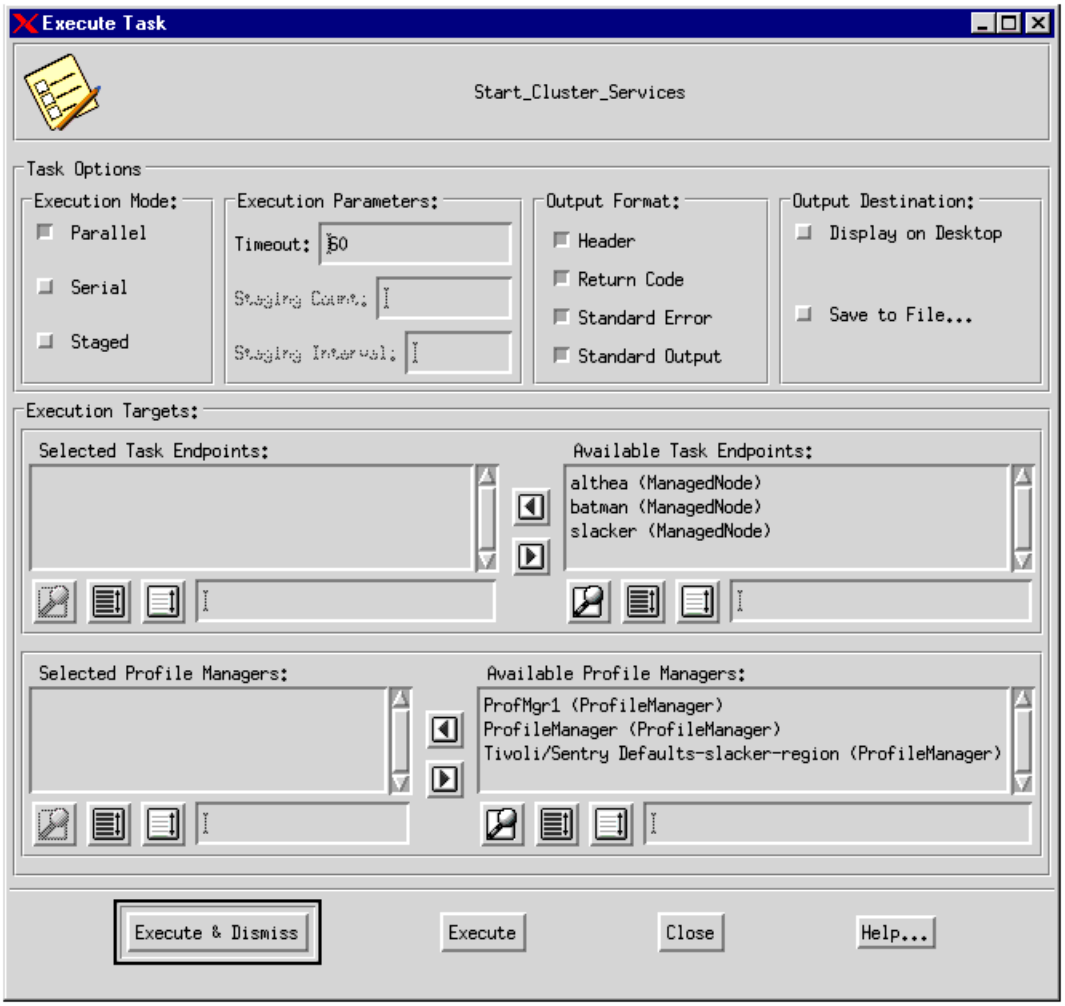
1. Select a task from the Task Library.
2. Specify appropriate Task Options.
3. Specify appropriate Task Arguments.
4. Execute the task.

These steps are detailed in the following sections.

Starting and Stopping Cluster Services via Tivoli

To configure starting and stopping of cluster services on specified nodes, perform the following steps:

1. From the HACMP Policy Region window, select the Cluster Services Task Library.
2. From the Cluster Services Task Library, select the task (Start_Cluster_Services or Stop_Cluster_Services) you want to perform. The Execute Task panel opens. (The panel for Start_Cluster_Services is shown here.)



Execute Task Panel for Starting Cluster Services

3. In the Execute Task panel (the panel for Start Cluster Services is shown below), set the appropriate Task Options for this task.

Notes on Task Options

- Select Display on Desktop if you want to see detailed output for each task immediately, or Save to File if you do *not* need to view the output now. (Even if you choose to view the display on the desktop, you will have the option to save it to a file later.)
- Task Description or Help buttons in Tivoli panels do *not* provide any information on HACMP-specific functionality.
- You will probably need to increase the Timeout parameter from its default of 60 seconds. Most tasks take longer than this to execute. For resource group tasks, factors such as the number of resource groups specified can cause the task to take as long as 10 minutes.

If the timeout period is set too short, then when you execute the task (from the Configure Task Arguments window), a message will appear to notify you when the timeout has expired. The event may have still completed successfully, but the detailed output will *not* appear.

- Specify the nodes on which you want to perform the task by moving the node names to the Selected Task Endpoints window. Click on the left-facing arrow to move selected node names from the Available list to the Selected list, and the right-facing arrow to move node names out of the Available list.)

  (Notice that the TMR node appears in the Available list. You are allowed to select it, but since it is *not* a cluster node, no action will be taken on that node. A message informs you of this.)
- The Selected Profile Managers and Available Profile Managers in the lower section of the panel are *not* used for HACMP monitoring.

4. After setting all necessary Task Options, click the Execute & Dismiss button.

   The Configure Task Arguments panel opens. (The panel for Start_Cluster_Services is shown here.)

Configure Task Arguments Panel for Starting Cluster Services

5. Configure the task arguments for starting or stopping cluster services, as you would when starting or stopping cluster services using SMIT.

   Note: Task Description or Help buttons in Tivoli panels do *not* provide any information on HACMP-specific functionality.

6. After setting all task arguments for a task, click the Set and Execute button. As the task executes, details appear in the display if you selected Display on Desktop. (See the note about the Timeout parameter under Step 3 above.)

The cluster will now start or stop according to the parameters you have set.

Bringing a Resource Group Online or Offline

Note: Resource groups are displayed in Tivoli as type "unknown."

To configure bringing a resource group online or offline:

1. Follow steps one through four under Starting and Stopping Cluster Services via Tivoli to select the task and configure the task options. The **Task Arguments** panel displays.
2. Select one or more resource groups and then configure the necessary task arguments for bringing a resource group online or offline, just as you would when performing this task using SMIT. (For more information on the task arguments for bringing resource groups online or offline, see Chapter 15: Managing Resource Groups in a Cluster.)

   Note: If you do *not* have admin level privileges, you are *not* allowed to select a resource group from the list. Instead, you see a message informing you that you have insufficient permissions.

3. After setting all task arguments for a task, click the Set and Execute button. As the task executes, details appear in the display if you selected Display on Desktop. (See the note about the Timeout parameter under Notes on Task Options above.)

The specified resource group(s) will now be brought online or offline as specified.

Moving a Resource Group

To configure that a resource group moves:

1. Follow steps one through four under Starting and Stopping Cluster Services via Tivoli to select the task and configure the task options. The **Task Arguments** panel displays.
2. Configure the necessary task arguments for moving a resource group, just as you would when performing this task through SMIT. (For more information on the task arguments for bringing resource groups online or offline, see Requirements before Migrating a Resource Group in Chapter 15: Managing Resource Groups in a Cluster.)

   Note: If you do *not* have admin level privileges, you are *not* allowed to select a resource group from the list. Instead, you see a message informing you that you have insufficient permissions.

3. After setting all task arguments for a task, click the Set and Execute button.

   As the task executes, details appear in the display if you selected Display on Desktop. (See the note about the Timeout parameter under Notes on Task Options above.)

The resource group will now be moved as specified.

Uninstalling HACMP-Related Files from Tivoli

To discontinue cluster monitoring or administration through Tivoli, you must perform the following steps to delete the HACMP-specific information from Tivoli:

1. Run an uninstall through the SMIT interface, uninstalling the three **hativoli** filesets on all cluster nodes and the TMR.
2. If it is *not* already running, invoke Tivoli on the TMR by entering these commands:
   - **. /etc/Tivoli/setup_env.sh**
   - **tivoli**
3. In the Policy Region for the cluster, select the **Modify HATivoli Properties** task library. A window appears containing task icons.
4. Select **Edit > Select All** to select all tasks, and then **Edit > Delete** to delete. The Operations Status window at the left shows the progress of the deletions.
5. Return to the Policy Region window and delete the **Modify HaTivoli Properties** icon.
6. Repeat steps 4 through 6 for the **Cluster Services** task library.
7. Open the Profile Manager.
8. Select **Edit > Profiles > Select All** to select all HACMP Indicators.
9. Select **Edit > Profiles > Delete** to delete the Indicators.
10. Unsubscribe the cluster nodes from the Profile Manager:
    a. In the Profile Manager window, select Subscribers.
    b. Highlight each HACMP node on the left, and click to move it to the right side.
    c. Click **Set & Close** to unsubscribe the nodes.

Monitoring Clusters with clstat

HACMP provides the **/usr/es/sbin/cluster/clstat** utility for monitoring a cluster and its components. The **clinfo** daemon must be running on the local node for this utility to work properly.

The **clstat** utility reports on the cluster components as follows:

- Cluster: cluster number (system-assigned); cluster state (up or down); cluster substate (stable, or unstable).
- Nodes: How many, and the state of each node (up, down, joining, leaving, or reconfiguring).

  For each node, **clstat** displays the IP label and IP address of each network interface attached to each node, and whether that interface is up or down. **clstat** does *not* display multiple IP labels on one network interface, as in networks with aliases.

  For each node, **clstat** displays service IP labels for serial networks and whether they are up or down.

Note: By default, **clstat** does *not* display whether the service IP labels for serial networks are down. Use **clstat -s** to display service IP labels on serial networks that are currently down.

For each node, **clstat** displays the states of any resource groups (per node): online or offline.

See the **clstat** man page for additional information.

The **/usr/es/sbin/cluster/clstat** utility runs on both ASCII and X Window Display clients in either single-cluster or multi-cluster mode. The client display automatically corresponds to the capability of the system. For example, if you run **clstat** on an X Window client, a graphical display appears; however, you can run an ASCII display on an X-capable machine by specifying the **-a** flag.

Viewing clstat with WebSMIT

With HACMP 5.4, you can use WebSMIT to:

- Display detailed cluster information.
- Navigate and view the status of the running cluster
- Configure and manage the cluster
- View graphical displays of sites, networks, nodes and resource group dependencies.

For more information on installing and configuring WebSMIT, see the *Installation Guide*. For more information on using WebSMIT, see Chapter 2: Administering a Cluster Using WebSMIT.

Viewing clstat in ASCII Display Mode

In ASCII display mode, you have the option of viewing status for a single cluster or multiple clusters. You can also use the **-o** option to save a single snapshot of the **clstat** output in a **cron** job.

Single-Cluster ASCII Display Mode

In single-cluster ASCII display mode, the **clstat** utility displays information about only one cluster. To invoke the **clstat** utility in single-cluster (non-interactive) mode, enter:

```
/usr/es/sbin/cluster/clstat
```

A panel similar to the following appears:

```
clstat - HACMP Cluster Status Monitor
               -------------------------------------

Cluster: myctestcluster (1044370190)
Tue Mar 11 14:19:50 EST 2004
               State: UP               Nodes: 2
               SubState: STABLE

        Node: holmes            State: UP
           Interface: holmes_en1svc (0)         Address: 192.168.90.40
                                                State:   UP
           Resource Group: econrg1                      State:  online

        Node: u853              State: UP
           Interface: u853_en1svc (0)           Address: 192.168.90.50
                                                State:   UP
           Resource Group: econrg1                      State:  online

***************** f/forward, b/back, r/refresh, q/quit ***************
```

clstat Single-Cluster ASCII Display Mode

The cluster information displayed shows the cluster ID and name. (Note that HACMP assigns the cluster ID number; this is *not* user-defined.) In this example, the cluster is up and has two nodes, both of which are up. Each node has one network interface. Note that the *forward* and *back* menu options apply when more than one page of information is available to display.

If more than one cluster exists when you run the **clstat** command, the utility notifies you of this fact and requests that you retry the command specifying one of the following options:

```
clstat [-c cluster ID] [-n name][ -r seconds] [-i] [-a] [-o] [-s]
```

where:

-c *cluster ID*	Displays information about the cluster with the specified ID if that cluster is active (HACMP generates this number). This option can*not* be used with the **-n** option. If the cluster is *not* available, the **clstat** utility continues looking for it until it is found or until the program is canceled. Note that this option can*not* be used if the **-i** option (for multi-cluster mode) is used.
-n *name*	The cluster name. This option can*not* be used with the **-c** option
-r *seconds*	Updates the cluster status display at the specified number of seconds. The default is 1 second; however, the display is updated only if the cluster state changes.
-i	Displays information about clusters interactively. Only valid when running **clstat** in ASCII mode.
-a	Causes **clstat** to display in ASCII mode.
-o	(once) Provides a single snapshot of the cluster state and exits. This flag can be used to run **clstat** out of a **cron** job. Must be run with the **-a** option; ignores **-i** or **-r** flags.

-s	Displays service labels for serial networks and their state (up or down).

To see cluster information about a specific cluster, enter:

```
clstat [-n name]
```

Multi-Cluster ASCII Display Mode

The multi-cluster (interactive) mode lets you monitor all clusters that Clinfo can access from the list of active service IP labels or addresses found in the **/usr/es/sbin/cluster/etc/clhosts** file. In multi-cluster mode, the **clstat** utility displays this list of recognized clusters and their IDs, allowing you to select a specific cluster to monitor. Multi-cluster mode requires that you use the **-i** flag when invoking the **clstat** utility. To invoke the **clstat** utility in multi-cluster mode, enter:

```
/use/es/sbin/cluster/clstat -i
```

where the **-i** indicates multi-cluster (interactive) ASCII mode. A panel similar to the following appears.

```
              clstat - HACMP for AIX Cluster Status Monitor
              ------------------------------------------------

Number of clusters active: 1

                ID      Name            State

              777     ibm_26c            UP

Select an option:

              # - the Cluster ID                      x- quit
```

clstat Multi-Cluster Mode Menu

This panel displays the ID, name, and state of each active cluster accessible by the local node. You can either select a cluster to see detailed information, or quit the **clstat** utility.

When you enter a cluster name, a panel appears similar to the one that follows.

```
              clstat - HACMP for AIX Cluster Status Monitor
                ---------------------------------------------

Cluster: ibm_26c        (777)           Thu Jul  9 18:35:46 EDT 2002
                State: UP Nodes: 2
                SubState: STABLE
        Node: poseidon          State: UP
           Interface: poseidon-enboot (0)       Address: 140.186.70.106
                                                State:   UP

        Node: venus             State: UP
           Interface: venus-enboot (0)          Address: 140.186.70.107
                                                State:   UP

       Resource Group: rot                    State: online
       Resource Gropu: rg1                    State: online
****************** f/forward, b/back, r/refresh, q/quit *************
```

clstat Multi-Cluster ASCII Display Mode

After viewing this panel, press **q** to exit the display. The multi-cluster mode returns you to the cluster list so you can select a different cluster. Note that you can use all menu options displayed. The *forward* and *back* options allow you to scroll through displays of active clusters without returning to the previous panel.

Viewing clstat in X Window System Display Mode

When you start the **/usr/es/sbin/cluster/clstat** utility on a node capable of displaying X Window System applications, the **clstat** utility displays its graphical interface if the client's DISPLAY environment variable is set to the value of the X server's node address.

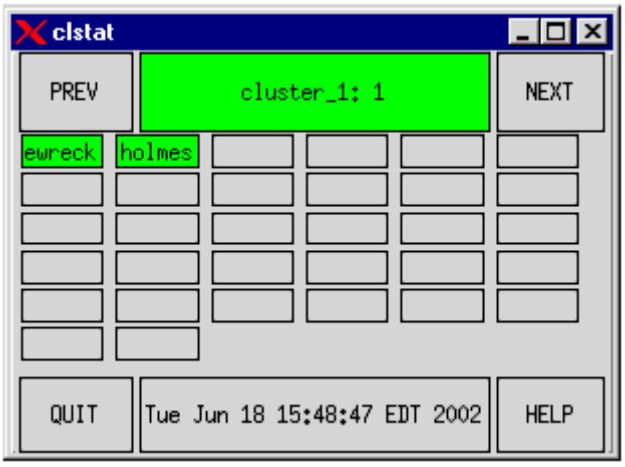
To invoke the **clstat** utility X Window System display, enter the **clstat** command:

```
/usr/es/sbin/cluster/clstat [-n name][-c Id][ -r #][-D debug_level][-s]
```

where:

-n *name*	The cluster name. This option can*not* be used with the **-c** option.
-c *ID*	Displays information about the cluster with the specified ID if that cluster is active. This option can*not* be used with the **-n** option.
-r #	The interval at which the **clstat** utility updates the display. For the graphical interface, this value is interpreted in tenths of seconds. By default, **clstat** updates the display every 0.10 seconds.
-D *debug_level*	The level of debugging to be performed. The levels range from 1 to 10 in increasing amounts of information. The default (0) turns debugging off.
-s	Displays service labels for serial networks and their state (up or down).

The **clstat** utility graphical interface uses windows to represent cluster nodes, as in the figure shown here:



clstat X Window System Display

The middle box in the top row indicates the cluster name and ID. If the cluster is stable, this box appears green. If the cluster destabilizes for any reason, this box changes to red.

The large boxes in other rows represent nodes. A node name appears in a box for each active node in the cluster. You can see up to sixteen nodes per cluster. Nodes that are up are shown in green, nodes that are down are shown in red, nodes that are joining or leaving the cluster are shown in yellow (topology changes), and nodes that are undefined are shown in the background color. Colors are configured in the **xclstat** X Window resource file in the **/usr/es/sbin/cluster/samples/clstat** directory.

On a monochrome display, gray shading represents the colors as follows:

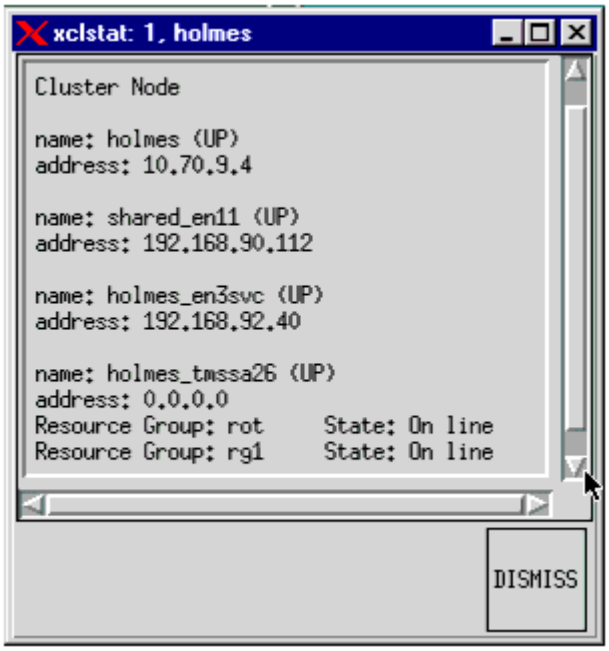
red	dark gray
yellow	gray
green	light gray

Five buttons are available on the **clstat** display:

PREV	Displays the previous cluster (loops from end to start).
NEXT	Displays the next cluster (loops from start to end).
cluster:ID	The refresh bar. Pressing this bar updates the status display.
QUIT	Cancels the **clstat** utility.
HELP	Displays help information.

Viewing Network Interface and Resource Group Information in an X Window Display

To view information about network interfaces and resource groups for a node, click mouse button 1 on the appropriate node box in the **clstat** display. A pop-up window similar to the following appears. The title in the example shows that you are viewing node *holmes* in *cluster_1*.



clstat Node Information Display

clstat displays only the state (online or offline) of resource groups.

Click on the DISMISS button to close the pop-up window and to return to the **clstat** display window. Do *not* use the **Close** option in the pull-down menu in the upper left corner of the window to close this display; it terminates the **clstat** utility.

Viewing clstat with a Web Browser

With an appropriately configured Web server, you can view **clstat** in a Web browser on any machine that can connect to the cluster node (a node with both a Web server and Clinfo running on it). Viewing **clstat** through a Web browser allows you to see status for all of your clusters on one panel, using hyperlinks or the scroll bar to view details for each cluster.

When you install HACMP, an executable file called **clstat.cgi** is installed in the same directory (**/usr/es/sbin/cluster/**) as the **clstat** and **xclstat** files. When run, **clstat.cgi** provides a CGI interface that allows cluster status output to be formatted in HTML and viewed in a Web browser.

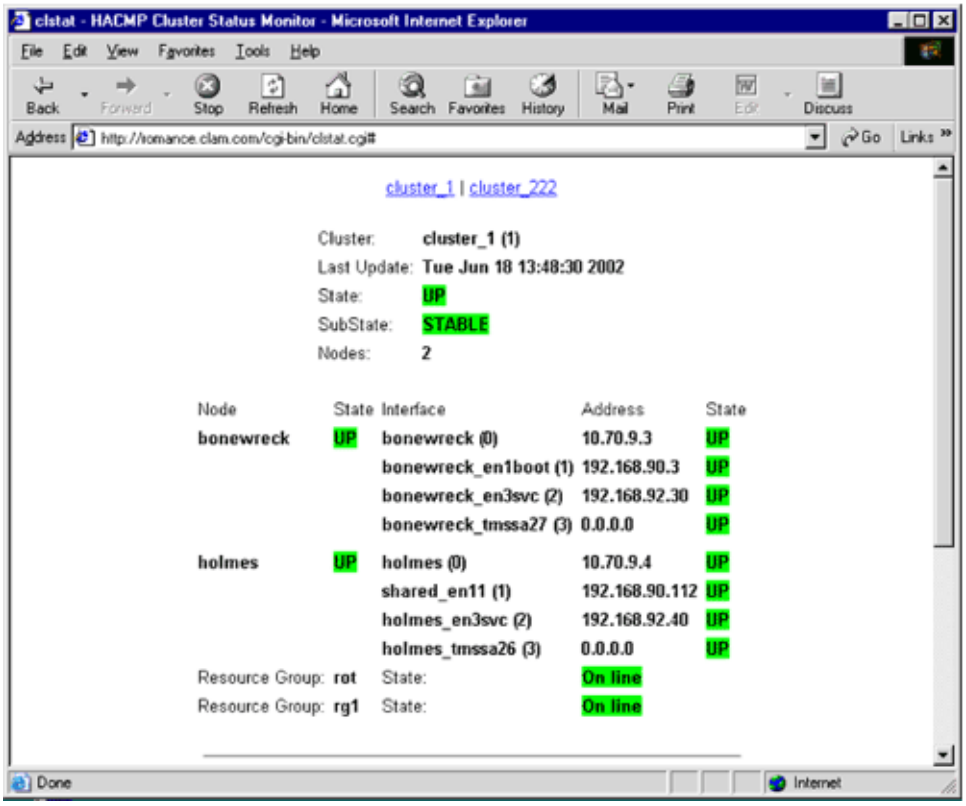
This feature supports the following browsers:

- Mozilla 1.7.3 for AIX and FireFox 1.0.6
- Internet Explorer, version 6.0.

Browser Display

The **clstat** HACMP Cluster Status Monitor displays the **clstat** output for all clusters from the list of active service IP labels or addresses found in the **/usr/es/sbin/cluster/etc/clhosts** file.

The example below shows **clstat** monitoring two clusters, *cluster_1* and *cluster_222*. The browser window displays the status information for one of the clusters, *cluster_1*. To display the other cluster, click the hyperlink for *cluster_222* at the top of the display or scroll down to find it.



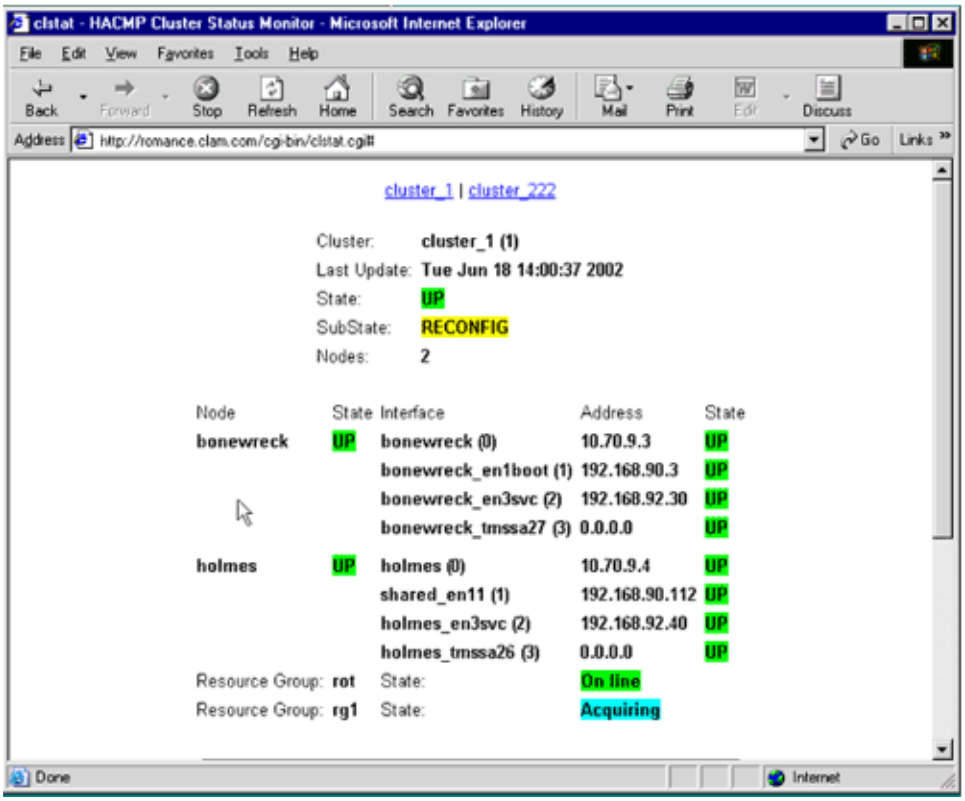
clstat Web Browser Display

The web browser display contains the same types of cluster status information as the ASCII or X Window displays, reorganized and color-coded for easier viewing.

The view automatically refreshes every 30 seconds to display current cluster status.

Note: After an automatic or manual refresh, the view should be retained; that is, the browser window should continue to display the cluster that was last clicked on before the refresh. In Internet Explorer 5.5 only, however, the refresh action causes a return to the top of the display.

In the following example, one of the resource groups is coming online and the cluster is therefore in a reconfiguration substate:



clstat Browser Display Showing a Resource Group in Acquiring State

Note: When a cluster resource group goes offline, it can no longer be displayed by **clstat**. No information about that resource group appears until it is being reacquired or online.

Configuring Web Server Access to clstat.cgi

To view the **clstat** display through a web browser, you must have a web server installed on a machine where Clinfo is running and able to gather cluster information. This could be a client node as well as a server node. The **clstat.cgi** program works with any web server that supports the CGI standard, which includes most currently available web servers for AIX 5L. For instance, you might use the IBM HTTP Server, which is included on the Expansion Pack CD for AIX 5L.

Full instructions for installing and configuring a web server are *not* included here. Please refer to the web server documentation or consult your web administrator if you need additional help.

The following steps complete the configuration of web server access to **clstat.cgi** using the IBM HTTP Server with its default configuration. The directories and URL you use for your server and configuration may vary.

1. Move or copy **clstat.cgi** to the cgi-bin or script directory of the web server, for instance the default HTTP Server directory **/usr/HTTPserver/cgi-bin**.
2. Verify that the **clstat.cgi** file still has appropriate permissions (that is, the file is executable by the user `nobody`).

3. You can now view cluster status using a web browser by typing in a URL of the following format:

```
http://<hostname or IP label of the web server node>/cgi-bin/clstat.cgi
```

Note: Although you can change the name of the CGI directory, do *not* rename the **clstat.cgi** file.

Changing the clstat.cgi Refresh Interval

You can change the default **clstat.cgi** refresh interval by specifying the CLSTAT_CGI_REFRESH environment variable in the **/etc/environment** file on the node serving the web page. Setting the CLSTAT_CGI_REFRESH environment variable (in seconds) overrides the default setting.

For example, to change the refresh interval to 15 seconds from the default setting, add the following to the **/etc/environment** file:

```
# change the clstat.cgi refresh interval to 15 seconds; 30 seconds is
the default

CLSTAT_CGI_REFRESH=15
```

Security

Because **clstat.cgi** is *not* run as root, there should be no immediate security threat of users gaining unauthorized access to HACMP by accessing **clstat.cgi** from the web server.

Some administrators may wish to restrict access to **clstat.cgi** from the web server and can use methods built in to the web server to prevent access, such as password authentication or IP address blocking. HACMP does *not* provide any specific means of access restriction to **clstat.cgi**.

Monitoring Applications

HACMP uses monitors to check if the application is running before starting the application, avoiding startup of an undesired second instance of the application. HACMP also monitors specified applications and attempts to restart them upon detecting process death or application failure.

Application monitoring works in one of two ways:

- *Process application monitoring* detects the termination of one or more processes of an application, using RSCT Resource Monitoring and Control (RMC).
- *Custom application monitoring* checks the health of an application with a custom monitor method at user-specified polling intervals.

HACMP uses monitors to check if the application is running before starting the application. You can configure multiple application monitors and associate them with one or more application servers. You can assign each monitor a unique name in SMIT.

By supporting multiple monitors per application, HACMP can support more complex configurations. For example, you can configure one monitor for each instance of an Oracle parallel server in use. Or, you can configure a custom monitor to check the health of the database along with a process termination monitor to instantly detect termination of the database process.

Process monitoring is easier to set up, as it uses the built-in monitoring capability provided by RSCT and requires no custom scripts; however, it may *not* be an appropriate option for all applications. User-defined monitoring can monitor more subtle aspects of an application's performance and is more customizable, but it takes more planning, as you must create the custom scripts.

In either case, when a problem is detected by the monitor, HACMP attempts to restart the application on the current node and continues the attempts until a specified retry count is exhausted. When an application can*not* be restarted within this retry count, HACMP takes one of two actions, which you specify when configuring the application monitor:

- Choosing **fallover** causes the resource group containing the application to fall over to the node with the next highest priority according to the resource policy.
- Choosing **notify** causes HACMP to generate a server_down event to inform the cluster of the failure.

When you configure an application monitor, you use the SMIT interface to specify which application is to be monitored and then define various parameters such as time intervals, retry counts, and action to be taken in the event the application can*not* be restarted. You control the application restart process through the Notify Method, Cleanup Method, and Restart Method SMIT fields, and by adding pre- and post-event scripts to any of the failure action or restart events you select.

You can temporarily suspend and then resume an application monitor in order to perform cluster maintenance.

When an application monitor is defined, each node's Configuration Database contains the names of monitored applications and their configuration data. This data is propagated to all nodes during cluster synchronization, and is backed up when a cluster snapshot is created. The cluster verification ensures that any user-specified methods exist and are executable on all nodes.

Note: If you specify the **fallover** option, which may cause a resource group to migrate from its original node, even when the highest priority node is up, the resource group may remain offline. Unless you bring the resource group online manually, it could remain in an inactive state. See Chapter 15: Managing Resource Groups in a Cluster, for more information.

A Note on Application Monitors

Application monitors configurable in HACMP are a critical piece of the HACMP cluster configuration; they enable HACMP to keep applications highly available. When HACMP starts an application server on a node, it uses a monitor that you configure to check if an application is already running to avoid starting two instances of the application. HACMP also periodically manages the application using the monitor that you configure to make sure that the application is up and running.

An erroneous application monitor may *not* detect a failed application. As a result, HACMP would *not* recover it or may erroneously detect an application as failed, which may cause HACMP to move the application to a takeover node, resulting in unnecessary downtime. For example, a custom monitor that uses an **sql** command to query a database to detect whether it is functional may *not* respond that the database process is running on the local node so this is *not* sufficient for use with HACMP.

If you plan on starting the cluster services with an option of **Manage Resources > Manually**, or stopping the cluster services without stopping the applications, HACMP relies on configured application monitors to determine whether to start the application on the node or *not*.

To summarize, we highly recommend properly configured and tested application monitors for all applications that you want to keep highly available with the use of HACMP. During verification, HACMP issues a warning if an application monitor is *not* configured.

For complete information on configuring application monitoring, see Configuring Multiple Application Monitors in Chapter 4: Configuring HACMP Cluster Topology and Resources (Extended).

Displaying an Application-Centric Cluster View

You can use either WebSMIT or the ASCII version of SMIT to view a cluster application.

To show a cluster application in SMIT:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Resource Configuration > HACMP Extended Resources Configuration > Configure HACMP Applications > Show Cluster Applications** and press Enter.

   SMIT displays the list of applications.
3. Select the application to show from the list.

   SMIT displays the application with its related components.

To show current resource group and application state, select **HACMP Resource Group and Application Management > Show Current Resource Group and Application State**. This panel displays the current states of applications and resource groups for each resource group.

- For non-concurrent groups, HACMP shows only the node on which they are online and the applications state on this node
- For concurrent groups, HACMP shows ALL nodes on which they are online and the applications states on the nodes
- For groups that are offline on all nodes, only the application states are displayed, node names are not listed.

The SMIT panel **Show All Resources by Node or Resource Group** has an option linking you to the **Show Current Resource Group and Application State**, described above.

Starting with HACMP 5.4, WebSMIT presents the application-centric information in the Navigation frame **Resource Groups View** tab. For more information, see Chapter 2: Administering a Cluster Using WebSMIT.

Measuring Application Availability

You can use the Application Availability Analysis tool to measure the amount of time that any of your applications is available. The HACMP software collects and logs the following information in time-stamped format:

- An application starts, stops, or fails
- A node fails, is shut down, or comes online (or cluster services are started or shut down)
- A resource group is taken offline or moved
- Application monitoring is suspended or resumed.

Using SMIT, you can select a time period and the tool will display uptime and downtime statistics for a given application during that period. The tool displays:

- Percentage of uptime
- Amount of uptime
- Longest period of uptime
- Percentage of downtime
- Amount of downtime
- Longest period of downtime
- Percentage of time application monitoring was suspended.

All nodes must be available when you run the tool to display the uptime and downtime statistics. Clocks on all nodes must be synchronized in order to get accurate readings.

The Application Availability Analysis tool treats an application that is part of a concurrent resource group as available as long as the application is running on any of the nodes in the cluster. Only when the application has gone offline on all nodes in the cluster will the Application Availability Analysis tool consider the application as unavailable.

The Application Availability Analysis tool reports application availability from the HACMP cluster infrastructure's point of view. It can analyze only those applications that have been properly configured so they will be managed by the HACMP software. (See the Appendix on Applications and HACMP in the *Planning Guide* for details on how to set up a highly available application with HACMP.)

When using the Application Availability Analysis tool, keep in mind that the statistics shown in the report reflect the availability of the HACMP application server, resource group, and (if configured) the application monitor that represent your application to HACMP.

The Application Availability Analysis tool can*not* detect availability from an end user's point of view. For example, assume that you have configured a client-server application so that HACMP manages the server, and, after the server was brought online, a network outage severed the connection between the end user clients and the server. The end users would view this as an application outage because their client software could *not* connect to the server, but HACMP would *not* detect it, because the server it was managing did *not* go offline. As a result, the Application Availability Analysis tool would *not* report a period of downtime in this scenario.

Planning and Configuring for Measuring Application Availability

If you have application servers defined, the Application Availability Analysis Tool automatically keeps the statistics for those applications.

In addition to using the Application Availability Analysis Tool, you can also configure Application Monitoring to monitor each application server's status. You can define either a Process Application Monitor or a Custom Application Monitor. (See the preceding section on Monitoring Applications for details.)

If you configure Application Monitoring solely for the purpose of checking on uptime status and do *not* want the Application Monitoring feature to automatically restart or move applications, you should set the **Action on Application Failure** parameter to just **Notify** and set the **Restart Count** to zero. (The default is three.)

Ensure that there is adequate space for the **clavan.log** file on the filesystem on which it is being written. Disk storage usage is a function of node and application stability (*not* availability), that is, of the number (*not* duration) of node or application failures in a given time period. Roughly speaking, the application availability analysis tool will use 150 bytes of disk storage per outage. For example, on a node that fails once per week and has one application running on it, where that application never fails on its own, this feature uses about 150 bytes of disk storage usage per week.

Whenever verification runs, it determines whether there is enough space for the log on all nodes in the cluster.

Configuring and Using the Application Availability Analysis Tool

To use SMIT to check on a given application over a certain time period:

1. Enter `smit hacmp`
2. In SMIT, select **System Management (C-SPOC)** > **Resource Group and Application Management > Application Availability Analysis** and press Enter.
3. Select an application. Press F4 to see the list of configured applications.
4. Fill in the fields as follows:

Application Name	Application you selected to monitor.
Begin analysis on year (1970-2038) **month (01-12)** **day (1-31)**	Enter 2006 for the year 2006, and so on.
Begin analysis at hour (00-23) **minutes (00-59)** **seconds (00-59)**	
End analysis on year (1970-2038) **month (01-12)** **day (1-31)**	

End analysis at hour (00-23)

minutes (00-59)

seconds (00-59)

5. Press Enter. The application availability report is displayed as shown in the sample below.

```
               COMMAND STATUS
Command: OK    stdout: yes    stderr: no
Before command completion, additional instructions may appear below.

Application: myapp

Analysis begins: Monday, 1-May-2002, 14:30
Analysis ends: Friday, 5-May-2002, 14:30

Total time:    5 days, 0 hours, 0 minutes, 0 seconds

Uptime:
Amount:        4 days, 23 hours, 0 minutes, 0 seconds
Percentage:    99.16 %
Longest period:     4 days, 23 hours, 0 minutes, 0 seconds

Downtime:
Amount:        0 days, 0 hours, 45 minutes, 0 seconds
Percentage:    00.62 %
Longest period:     0 days, 0 hours, 45 minutes, 0 seconds
```

If the utility encounters an error in gathering or analyzing the data, it displays one or more error messages in a **Command Status** panel.

Reading the clavan.log File

The application availability analysis log records are stored in the **clavan.log** file. The default directory for this log file is **/var/adm**. You can change the directory using the **System Management C-SPOC > HACMP Log Viewing and Management > Change/Show a HACMP Log Directory** SMIT panel. Each node has its own instance of the file. You can look at the logs at any time to get the uptime information for your applications.

Note: If you redirect the log, remember it is a cumulative file. Its usefulness for statistical information and analysis will be affected if you do *not* keep the information in one place.

clavan.log file format

The **clavan.log** file format is described here.

```
Purpose
Records the state transitions of applications managed by HACMP.

Description
The clavan.log file keeps track of when each application that is managed
by HACMP is started or stopped and when the node stops on which
an application is running.  By collecting the records in the
clavan.log file from every node in the cluster, a utility program
can determine how long each application has been up, as well as
compute other statistics describing application availability time.

Each record in the clavan.log file consists of a single line.
```

```
Each line contains a fixed portion and a variable portion:

AAA: Ddd Mmm DD hh:mm:ss:YYYY: mnemonic:[data]:[data]: <variable
portion>

Where:         is:
------        ----
AAA           a keyword
Ddd            the 3-letter abbreviation for the day of the week
YYYY          the 4-digit year
Mmm            The 3-letter abbreviation for month
DD             the 2-digit day of the month (01...31)
hh             the 2-digit hour of the day (00...23)
mm             the 2-digit minute within the hour (00...59)
ss             the 2-digit second within the minute (00...59)

variable portion: one of the following, as appropriate (note that umt
stands for Uptime Measurement Tool, the original name of this tool):
```

Mnemonic	Description	As used in clavan.log file
umtmonstart	monitor started	umtmonstart:monitor_name:node:
umtmonstop	monitor stopped	umtmonstop:monitor_name:node:
umtmonfail	monitor failed	umtmonfail:monitor_name:node:
umtmonsus	monitor suspended	umtmonsus:monitor_name:node:
umtmonres	monitor resumed	umtmonres:monitor_name:node:
umtappstart	application server started	umtappstart:app_server:node:
umtappstop	application server stopped	umtappstop:app_server:node:
umtrgonln	resource group online	umtrgonln:group:node:
umtrgoffln	resource group offline	umtrgoffln:group:node:
umtlastmod	file last modified	umtlastmod:date:node:
umtnodefail	node failed	umtnodefail:node:
umteventstart	cluster event started	umteventstart:event
[arguments]:		
umteventcomplete	cluster event completed	umteventcomplete:event
[arguments]:		

```
Implementation Specifics

None.

Files

/var/adm/clavan.log
This is the default file spec for this log file.
```

```
The directory can be changed with the "Change/Show a
HACMP Log Directory" SMIT panel (fast path =
"clusterlog_redir_menu")

Related Information

None.
```

Examples

The following example shows output for various types of information captured by the tool.

```
AAA: Thu Feb 21 15:27:59 2002: umteventstart:reconfig_resource_release:
Cluster event reconfig_resource_release started
AAA: Thu Feb 21 15:28:02 2002:
umteventcomplete:reconfig_resource_release: Cluster event
reconfig_resource_release completed
AAA: Thu Feb 21 15:28:15 2002: umteventstart:reconfig_resource_acquire:
Cluster event reconfig_resource_acquire started
AAA: Thu Feb 21 15:30:17 2002:
umteventcomplete:reconfig_resource_acquire: Cluster event
reconfig_resource_acquire completed
AAA: Thu Feb 21 15:30:17 2002: umteventstart:reconfig_resource_complete:
Cluster event reconfig_resource_complete started
AAA: Thu Feb 21 15:30:19 2002: umtappstart:umtappa2:titan: Application
umtappa2 started on node titan
AAA: Thu Feb 21 15:30:19 2002: umtrgonln:rota2:titan: Resource group
rota2 online on node titan
```

Notes

clavan.log file records are designed to be human-readable but also easily parsed. This means you can write your own analysis programs. The Application Availability Analysis tool is written in Perl and can be used as a reference for writing your own analysis program. The pathname of the tool is **/usr/es/sbin/cluster/utilities/clavan**.

Using Resource Groups Information Commands

In addition to using the HAView utility to monitor resource group status and location, as discussed earlier in this chapter, you can locate resource groups using the command line.

You can use the **/usr/es/sbin/cluster/utilities/clRGinfo** command to monitor resource group status and location. The command tells you the current location and if a node temporarily has the highest priority for this instance.

For the complete description and examples of the command usage see the **clRGinfo** section in Appendix A: Script Utilities in the *Troubleshooting Guide*.

Note: Alternatively, you can use the **clfindres command** instead of **clRGinfo**. **clfindres** is a link to **clRGinfo**. Only the root user can run the **clRGinfo** utility.

Using the clRGinfo Command

Running the **clRGinfo** command gives you a report on the location and state of one or more specified resource groups. The output of the command displays both the global state of the resource group as well as the special state of the resource group on a local node. A resource group can be in any one of the following states (if sites are configured, more states are possible):

- *Online.* The resource group is currently operating properly on one or more nodes in the cluster.
- *Offline*. The resource group is *not* operating in the cluster and is currently *not* in an error condition. Two particular reasons for an Offline state are displayed in these cases:
  - *OFFLINE Unmet Dependencies*
  - *OFFLINE User Requested*
- *Acquiring*. A resource group is currently coming up on one of the nodes in the cluster.
- *Releasing.* The resource group is in the process of being released from ownership by one node. Under normal conditions after being successfully released from a node the resource group's status changes to offline.
- *Error.* The resource group has reported an error condition. User interaction is required.
- *Unknown.* The resource group's current status can*not* be attained, possibly due to loss of communication; the fact that all nodes in the cluster are *not* up, or because a resource group dependency is *not* met (another resource group that depends on this resource group failed to be acquired first).

Resource Group States with Sites Defined

If sites are defined in the cluster, the resource group can be in one of the following states:

On the Primary Site	On the Secondary Site
ONLINE	ONLINE SECONDARY
OFFLINE	OFFLINE SECONDARY
ERROR	ERROR SECONDARY
UNMANAGED	UNMANAGED SECONDARY

Depending on the **Inter-Site Management Policy** defined in SMIT for a resource group, a particular resource group can be online on nodes in both primary and secondary sites. A resource group can also be online on nodes within one site and offline on nodes within another.

You can use the Resource Group Management utility, **clRGmove**, to move a resource group online, offline or to another node either within the boundaries of a particular site, or to the other site. For more information on moving resource groups with sites defined, see the section Migrating Resource Groups with Replicated Resources in Chapter 15: Managing Resource Groups in a Cluster.

Note: Only one instance of a resource group state exists for OFFLINE and ERROR states, be it on a primary or a secondary site. The **clRGinfo** command displays a node on a particular site, the state of a resource group, and a node that temporarily has the highest priority for this instance.

clRGinfo Command Syntax

The **clRGinfo -a** command provides information on what resource group movements take place during the current cluster event. For concurrent resource groups, it indicates on which nodes a resource group goes online or offline.

If **clRGinfo** can*not* communicate with the Cluster Manager on the local node, it attempts to find a cluster node with the Cluster Manager running, from which resource group information may be retrieved. If **clRGinfo** fails to find at least one node with the Cluster Manager running, HACMP displays an error message.

clRGinfo has the following syntax:

```
clRGinfo [-h][-v][-a][-s|-c][-p][-t][-d][groupname1] [groupname2] ...
```

Using **clRGinfo -a** in pre and post-event scripts is recommended, especially in HACMP clusters with dependent resource groups. When HACMP processes dependent resource groups, multiple resource groups can be moved at once with the **rg_move** event.

- Use **clRGinfo -t** to query the Cluster Manager on the local node only. This option displays the resource groups on the local node with the settling time and the delayed fallback timer settings, if they were set for a resource group.

clRGinfo Command Sample Outputs

The following examples show the output of the **clRGinfo** command.

The **clRGinfo -a** command lets you know the pre-event location and the post-event location of a particular resource group, as in the following examples:

Note: **clRGinfo - a** provides meaningful output *only* if you run it while a cluster event is being processed.

- In this example, the resource group A is moving from the offline state to the online state on node B. The pre-event location is left blank, the post-event location is Node B:

```
:rg_move[112] /usr/es/sbin/cluster/utilities/clRGinfo -a
-----------------------------------------------------
Group Name Resource Group Movement
-----------------------------------------------------
rgA            PRIMARY=":nodeB"
```

- In this example, the resource group B is moving from Node B to the offline state. The pre-event location is node B, the post-event location is left blank:

```
:rg_move[112] /usr/es/sbin/cluster/utilities/clRGinfo -a
-----------------------------------------------------
Group Name Resource Group Movement
-----------------------------------------------------
rgB            PRIMARY="nodeB:"
```

- In this example, the resource group C is moving from Node A to Node B. The pre-event location is node A, the post-event location is node B:

```
:rg_move[112] /usr/es/sbin/cluster/utilities/clRGinfo -a
-----------------------------------------------------------
Group Name Resource Group Movement
-----------------------------------------------------------
rgC            PRIMARY="nodeA:nodeB"
```

- In this example with sites, the primary instance of resource group C is moving from Node A to Node B, and the secondary instance stays on node C:

```
:rg_move[112] /usr/es/sbin/cluster/utilities/clRGinfo -a
-----------------------------------------------------------
Group Name Resource Group Movement
-----------------------------------------------------------
rgC            PRIMARY="nodeA:nodeB"
               SECONDARY=”nodeC:nodeC”
```

- With concurrent resource groups, the output indicates each node from which a resource group is moving online or offline. In the following example, both nodes release the resource group:

```
:rg_move[112] /usr/es/sbin/cluster/utilities/clRGinfo -a
-----------------------------------------------------------
Group Name Resource Group Movement
-----------------------------------------------------------
rgA    "nodeA:"
rgA    "nodeB:"
```

Because HACMP performs these calculations at event startup, this information will be available in pre-event scripts (such as a pre-event script to **node_up**), on all nodes in the cluster, regardless of whether the node where it is run takes any action on a particular resource group.

With this enhancement to **clRGinfo**, the specific behaviors of each resource group can be further tailored with pre- and post-event scripts.

The **clRGinfo -c|-s -p** command lists the output in a colon separated format, and indicates the node that is temporarily the highest priority node, if applicable.

Here is an example:

```
$ clRGinfo -s -p
$ /usr/es/sbin/cluster/utilities/clRGinfo -s
Group1:ONLINE:merry::ONLINE:OHN:FNPN:FBHPN:ignore: : : :ONLINE:
Group1:OFFLINE:samwise::OFFLINE:OHN:FNPN:FBHPN:ignore: : : :ONLINE:
Group2:ONLINE:merry::ONLINE:OAAN:BO:NFB:ignore: : : :ONLINE:
Group2:ONLINE:samwise::ONLINE:OAAN:BO:NFB:ignore: : : :ONLINE:
```

Note: The **-s** flag prints the output in the following order:

```
RGName:node:(empty):nodeState:startup:fallover:fallback: \
 intersite:nodePOL:POL_SEC: \
 fallbackTime:settlingTime: \
 globalState:siteName:sitePOL
```

where the resource group's startup fallover and fallback preferences are abbreviated as follows:

```
Resource group's startup policies:
    OHN: Online On Home Node Only
```

```
    OFAN: Online On First Available Node
    OUDP: Online Using Node Distribution Policy
    OAAN: Online On All Available Nodes
Resource group's fallover policies:
    FNPN: Fallover To Next Priority Node In The List
    FUDNP: Fallover Using Dynamic Node Priority
    BO: Bring Offline (On Error Node Only)
Resource group's fallback policies:
    FHPN: Fallback To Higher Priority Node In The List
    NFB: Never Fallback
Resource group's intersite policies:
    ignore: ignore
    OES: Online On Either Site
    OBS: Online Both Sites
    PPS: Prefer Primary Site
```

If an attribute is *not* available for a resource group, the command displays a colon and a blank instead of the attribute.

The **clRGinfo -p** command displays the node that temporarily has the highest priority for this instance as well as the state for the primary and secondary instances of the resource group. The command shows information about those resource groups whose locations were temporally changed because of user-requested rg_move events.

```
$ /usr/es/sbin/cluster/utilities/clRGinfo -p
here3!

Cluster Name: TestCluster

Resource Group Name: Parent
Primary instance(s):
The following node temporarily has the highest priority for this
instance:
user-requested rg_move performed on Wed Dec 31 19:00:00 1969

Node                          State
----------------------------- ---------------
node3@s2                    OFFLINE
node2@s1                    ONLINE
node1@s0                    OFFLINE

Resource Group Name: Child
Node                          State
----------------------------- ---------------
node3@s2                    ONLINE
node2@s1                    OFFLINE
node1@s0                    OFFLINE
```

The **clRGinfo -p -t** command displays the node that temporarily has the highest priority for this instance and a resource group's active timers:

```
/usr/es/sbin/cluster/utilities/clRGinfo -p -t
Cluster Name: MyTestCluster

Resource Group Name: Parent
Primary instance(s):
```

The following node temporarily has the highest priority for this instance:

node4, user-requested rg_move performed on Fri Jan 27 15:01:18 2006

```
Node           Primary State    Secondary StateDelayed Timers
------------------------------ --------------- -------------------
node1@siteA    OFFLINE          ONLINE SECONDARY
node2@siteA    OFFLINE          OFFLINE
node3@siteB    OFFLINE          OFFLINE
node4@siteB    ONLINE           OFFLINE

Resource Group Name: Child
Node           State            Delayed Timers
---------------------------- --------------- -------------------
node2          ONLINE
node1          OFFLINE
node4          OFFLINE
node3          OFFLINE
```

The **clRGinfo -v** command displays the resource group's startup, fallover and fallback preferences:

```
$ /usr/sbin/cluster/utilites/clRGinfo -v

Cluster Name: MyCluster
Resource Group Name: myResourceGroup

Startup Policy: Online On Home-Node Only
Fallover Policy: Fallover Using Dynamic Node Priority
Fallback Policy: Fallback To Higher Priority Node In The List
Site Policy: Ignore

Location       State
-----------------------------------------
nodeA          OFFLINE
nodeB          ONLINE
nodeC          ONLINE
```

Using the cldisp Command

The **/usr/es/sbin/cluster/utilities/cldisp** command provides the application-centric view of the cluster configuration. This utility can be used to display resource groups and their startup, fallover, and fallback policies.

To show cluster applications:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Resource Configuration > Configure HACMP Applications > Show Cluster Applications** and press Enter.

   SMIT displays the information as shown in the example:

   ```
   #############
   ```

```
APPLICATIONS
#############
   Cluster HAES_52_Test_Cluster_Cities provides the following
applications: Application_Server_1 Application_Server_NFS_10
      Application: Application_Server_1 State: {online}

Application 'Application_Server_NFS_10' belongs to a resource group
which is configured to run on all its nodes simultaneously. No
fallover will occur.
 This application is part of resource group 'Resource_Group_03'.
            The resource group policies:
               Startup: on all available nodes
              Fallover: bring offline on error node
               Fallback: never
         Nodes configured to provide Application_Server_1:
Node_Kiev_1{up} Node_
Minsk_2{up} Node_Moscow_3{up}
            Nodes currently providing Application_Server_1:
Node_Kiev_1{up} Node
_Minsk_2{up} Node_Moscow_3{up}
         Application_Server_1 is started by
/usr/user1/hacmp/local/ghn_start_4
         Application_Server_1 is stopped by
/usr/user1/hacmp/local/ghn_stop_4
         Resources associated with Application_Server_1:
            Concurrent Volume Groups:
Volume_Group_03
         No application monitors are configured for
Application_Server_1.

      Application: Application_Server_NFS_10 State: {online}
         This application is part of resource group
'Resource_Group_01'.
            The resource group policies:
               Startup: on home node only
               Fallover: to next priority node in the list
               Fallback: if higher priority node becomes available
         Nodes configured to provide Application_Server_NFS_10:
Node_Kiev_1{up}...
```

Here is an example of the text output from the **cldisp** command:

```
app1{online}
This application belongs to the resource group rg1.
Nodes configured to provide app1: unberto{up} lakin{up}
The node currently providing app1 is: unberto {up}
The node that will provide app1 if unberto fails is: lakin
app1 is started by /home/user1/bin/app1_start
app1 is stopped by /home/user1/bin/app1_stop
Resources associated with app1:
srv1(10.10.11.1){online}
Interfaces are configured to provide srv1:
lcl_unberto (en1-10.10.10.1) on unberto{up}
lcl_lakin (en2-10.10.10.2) on lakin{up}
Shared Volume Groups: NONE
Concurrent Volume Groups: NONE
Filesystems: NONE
AIX Fast Connect Services: NONE
Application monitor of app1: app1
Monitor: app1
Type: custom
Monitor method: /home/user1/bin/app1_monitor
Monitor interval: 30 seconds
Hung monitor signal: 9
Stabilization interval: 30 seconds
Retry count: 3 tries
```

```
Restart interval: 198 seconds
Failure action: notify
Notify method: /home/user1/bin/app1_monitor_notify
Cleanup method: /home/user1/bin/app1_stop
Restart method: /home/user1/bin/app1_start
```

Using HACMP Topology Information Commands

You can see the complete topology configuration using the **/usr/es/sbin/cluster/utilities/cltopinfo** command. See Appendix C: HACMP for AIX Commands for the complete syntax and examples with various flags. to organize the information by node, network, or network interface. The following example uses the basic command:

```
$ /usr/es/sbin/cluster/utilities/cltopinfo
Cluster Description of Cluster: FVT_mycluster
Cluster Security Level: Standard
There are 2 node(s) and 1 network(s) defined

NODE holmes:
        Network ether_ipat
              sherlock_en3svc_a1  192.168.97.50
                holmes_en1svc_a1      192.168.95.40
                holmes_en1svc   192.168.90.40

NODE sherlock:
        Network ether_ipat
              sherlock_en3svc_a1  192.168.97.50
                holmes_en1svc_a1        192.168.95.40
              sherlock_en1svc     192.168.90.50

Resource Group econrg1
        Behavior                concurrent
        Participating Nodes     holmes sherlock
```

Monitoring Cluster Services

After checking cluster, node, and network interface status, check the status of the HACMP and RSCT daemons on both nodes and clients.

Monitoring Cluster Services on a Node

Depending on what you need to know, you may access the following for information:

- View Management Information Base (MIB) in the hacmp.out file.
- Use SMIT to check the status of the following HACMP subsystems on a node:
  - Cluster Manager (**clstrmgrES**) subsystem
  - SNMP (**snmpd**) daemon.
  - Clinfo (**clinfoES**) Cluster Information subsystem.
- To view cluster services on a node, enter the fastpath `smit clshow`

  A panel similar to following appears.

```
                        COMMAND STATUS

Command: OK            stdout: yes           stderr: no

Before command completion, additional instructions may appear below.

Subsystem             Group            PID     Status
clstrmgrES            cluster          18524   active
clinfoES              cluster          15024   active
```

Monitoring Cluster Services on a Client

The only HACMP process that can run on a client is the Cluster Information (**clinfo**) daemon. (Not all clients run this daemon.) You can use the AIX 5L **lssrc** command with either the **-g cluster** or **-s clinfoES** arguments to check the status of the **clinfo** subsystem on a client. The output looks similar to the following:

```
Subsystem      Group   PID     Status

clinfoES       cluster 9843    active
```

You can also use the **ps** command and **grep** for “clinfo.” For example:

```
ps -aux | grep clinfoES
```

HACMP Log Files

HACMP writes the messages it generates to the system console and to several log files. Because each log file contains a different subset of the types of messages generated by HACMP, you can get different views of cluster status by viewing different log files. HACMP writes messages into the log files described below. For more information about these files see Chapter 2: Using Cluster Log Files in the *Troubleshooting Guide*.

The default locations of log files are used in this chapter. If you redirected any logs, check the appropriate location.

Note: If you redirect logs, they should be redirected to local filesystems and *not* to shared or NFS filesystems. Having logs on shared or NFS filesystems may cause problems if the filesystem needs to unmount during a fallover event. Redirecting logs to shared or NFS filesystems may also prevent cluster services from starting during node reintegration.

Size of /var Filesystem May Need to Be Increased

For each node in your cluster, verification requires from 500K to 4 MB of free space in the **/var** filesystem. HACMP stores, at most, four different copies of a node's verification data on a disk at a time:

- **/var/hacmp/clverify/current/<nodename>/*** contains logs from a current execution of cluster verification
- **/var/hacmp/clverify/pass/<nodename/*** contains logs from the last time verification passed

- **/var/hacmp/clverify/pass.prev/<nodename/*** contains logs from the second to last time verification passed
- /**var/hacmp/clverify/fail/<nodename>/*** contains information from the last time verification failed.

The **/var/hacmp/clverify/clverify.log[0-9]** log files typically consume 1-2 MB of disk space.

In addition, the standard security mechanism that runs the **clcomd** utility has the following requirements for the free space in the **/var** filesystem:

1) 20 MB, where:

- **/var/hacmp/clcomd/clcomd.log** requires 2MB
- **/var/hacmp/clcomd/clcomddiag.log** requires 18MB.

2) 1 MB x `n`, per node (where `n` is the number of nodes in the cluster) in the file **/var/hacmp/odmcache**.

To summarize, for a four-node cluster it is recommended to have at least 42MB of free space in the **/var** filesystem, where:

- 2MB should be free for writing the **clverify.log[0-9]** files
- 16MB (4MB per node) for writing the verification data from the nodes
- 20MB for writing the **clcomd** log information
- 4MB (1MB per node) for writing the ODMcache data.

/tmp/clinfo.debug File

Clinfo is typically installed on both client and server systems. Client systems (**cluster.es.client**) do *not* contain any HACMP ODMs (for example HACMPlogs) or utilities (for example clcycle) therefore the logging for Clinfo does *not* take advantage of the redirection or cycling.

The **/tmp/clinfo.debug** file records the output generated by the event scripts as they run. This information supplements and expands upon the information in the **/usr/var/hacmp/log** file.

The default bug level is 0 or **OFF.** You can change the log file name with the command **clinfo -l**. See the manpage for more information.

/tmp/clsmuxtrmgr.debug Log File

The **clsumxtrmgr.debug** is the smux peer function log file. The default is *no debugging.* You can toggle the smux peer tracing on and off using the AIX 5L System Resource Controller (SRC).

/tmp/hacmp.out File

The **/tmp/hacmp.out** file records the output generated by the event scripts as they execute. This information supplements and expands upon the information in the **/usr/es/adm/cluster.log** file. To receive verbose output, the **debug level** runtime parameter should be set to *high* (the default). For details on setting runtime parameters see Chapter 1: Troubleshooting HACMP Clusters in the *Troubleshooting Guide*.

Reported resource group acquisition failures (failures indicated by a non-zero exit code returned by a command) are tracked in **hacmp.out**, and a summary is written near the end of the **hacmp.out** listing for a top-level event.

Checking this log is important, since the **reconfig_too_long** console message is *not* evident in every case where a problem exists. Event summaries make it easier for you to check the **hacmp.out** file for errors. For more information about this log file and about how to get a quick view of several days' event summaries see the Understanding the hacmp.out Log File section and Displaying Compiled hacmp.out Event Summaries section in Chapter 2: Using Cluster Log Files in the *Troubleshooting Guide*.

In releases prior to HACMP 5.2, non-recoverable event script failures result in the **event_error** event being run on the cluster node where the failure occured. The remaining cluster nodes do *not* indicate the failure. With HACMP 5.2 and up, all cluster nodes run the **event_error** event if any node has a fatal error. All nodes log the error and call out the failing node name in the **hacmp.out** log file.

/tmp/clstrmgr.debug Log File

The **clstrmgr.debug** log file contains time-stamped, formatted messages generated by HACMP **clstrmgrES** activity. This file is typically used only by IBM support personnel. The **/usr/es/sbin/cluster/utilities/clgetesdbginfo** command collects all the cluster log files. IBM support may ask you to run this command.

/tmp/cspoc.log File

The **cpoc.log** file contains time-stamped, formatted messages generated by HACMP C-SPOC commands. The **/tmp/cspoc.log** file resides on the node from which you issue the C-SPOC command.

/tmp/emuhacmp.out File

The **/tmp/emuhacmp.out** file records the output generated by the event emulator scripts as they execute. The **/tmp/emuhacmp.out** file resides on the node from which the event emulator is invoked. You can use the environment variable EMUL_OUTPUT to specify another name and location for this file, but the format and information remains the same.

/usr/es/adm/cluster.log File

The **/usr/es/adm/cluster.log** file is the main HACMP log file. HACMP error messages and messages about HACMP-related events are appended to this log with the time and date at which they occurred.

/usr/es/sbin/cluster/history/cluster.mmddyyyy File

The **/usr/es/sbin/cluster/history/cluster.*mmddyyyy*** file contains time-stamped, formatted messages generated by HACMP scripts. The system creates a cluster history file whenever cluster events occur, identifying each file by the file name extension *mmddyyyy*, where *mm* indicates the month, *dd* indicates the day, and *yyyy* indicates the year.

While it is more likely that you will use these files during troubleshooting, you should occasionally look at them to get a more detailed idea of the activity within a cluster.

/var/adm/clavan.log File

The **clavan.log** file keeps track of when each application that is managed by HACMP is started or stopped and when the node stops on which an application is running. By collecting the records in the **clavan.log** file from every node in the cluster, a utility program can determine how long each application has been up, as well as compute other statistics describing application availability time.

/var/hacmp/clcomd/clcomd.log File

The **clcomd.log** file contains time-stamped, formatted messages generated by the HACMP Cluster Communication Daemon. This log file contains an entry for every connect request made to another node and the return status of the request.

For information on space requirements for this file and for the file described below, see the section Size of /var Filesystem May Need to Be Increased.

/var/hacmp/clcomd/clcomddiag.log File

The **clcomddiag.log** file contains time-stamped, formatted messages generated by the HACMP Communication daemon when tracing is turned on. This log file is typically used by IBM support personnel for troubleshooting.

/var/hacmp/clverify/clverify.log File

The **/var/hacmp/clverify/clverify.log** file contains verbose messages, output during verification. Cluster verification consists of a series of checks performed against various HACMP configurations. Each check attempts to detect either a cluster consistency issue or an error. The verification messages follow a common, standardized format, where feasible, indicating such information as the node(s), devices, and command in which the error occurred. See Chapter 7: Verifying and Synchronizing an HACMP Cluster for complete information.

For information on space requirements for this file, see the section Size of /var Filesystem May Need to Be Increased earlier in this chapter.

/var/hacmp/log/clutils.log File

The **/var/hacmp/log/clutils.log** file contains the results of the automatic verification that runs on one user-selectable HACMP cluster node once every 24 hours. When cluster verification completes on the selected cluster node, this node notifies the other cluster nodes with the following information:

- The name of the node where verification had been run
- The date and time of the last verification
- Results of the verification.

The **/var/hacmp/log/clutils.log** file also contains messages about any errors found and actions taken by HACMP for the following utilities:

- The HACMP File Collections utility
- The Two-Node Cluster Configuration Assistant
- The Cluster Test Tool

- The OLPW conversion tool.

/var/ha/log/grpsvcs.<filename> File

Contains time-stamped messages in ASCII format. These track the execution of internal activities of the **grpsvcs** daemon. IBM support personnel use this information for troubleshooting. The file gets trimmed regularly. Therefore please save it promptly if there is a chance you may need it.

/var/ha/log/topsvcs.<filename> File

The **/var/ha/log/topsvcs.<filename>** log file contains time-stamped messages in ASCII format. These track the execution of internal activities of the **topsvcs** daemon. IBM support personnel use this information for troubleshooting. The file gets trimmed regularly. Therefore please save it promptly if there is a chance you may need it.

/var/ha/log/grpglsm File

The **/var/ha/log/grpglsm** file tracks the execution of internal activities of the **grpglsm** daemon. IBM support personnel use this information for troubleshooting. The file gets trimmed regularly. Therefore please save it promptly if there is a chance you may need it.

Chapter 11: Managing Shared LVM Components

This chapter explains how to maintain AIX 5L Logical Volume Manager (LVM) components shared by nodes in an HACMP cluster and provides procedures for managing volume groups, filesystems, logical volumes, and physical volumes using the HACMP Cluster-Single Point of Control (C-SPOC) utility.

The C-SPOC utility simplifies maintenance of shared LVM components in clusters of up to 32 nodes. C-SPOC commands provide comparable functions in a cluster environment to the standard AIX 5L commands that work on a single node. By automating repetitive tasks, C-SPOC eliminates a potential source of errors, and speeds up the cluster maintenance process.

In SMIT, you access C-SPOC using the **System Management (C-SPOC)** menu.

Although you can also use AIX 5L on each node to do these procedures, using the C-SPOC utility ensures that all commands are executed in the proper order. For specific information on AIX 5L commands and SMIT panels, see your *AIX 5L System Management Guide*.

The main topics in this chapter include:

Overview

A key element of any HACMP cluster is the data used by the highly available applications. This data is stored on AIX 5L LVM entities. HACMP clusters use the capabilities of the LVM to make this data accessible to multiple nodes.

In an HACMP cluster, the following definitions are used:

- A *shared volume group* is a volume group that resides entirely on the external disks shared by cluster nodes.
- A *shared physical volume* is a disk that resides in a shared volume group.
- A *shared logical volume* is a logical volume that resides entirely in a shared volume group.
- A *shared filesystem* is a filesystem that resides entirely in a shared logical volume.

Common Maintenance Tasks

As a system administrator of an HACMP cluster, you may be called upon to perform any of the following LVM-related tasks:

- Creating a new shared volume group
- Extending, reducing, changing, or removing an existing volume group
- Creating a new shared logical volume
- Extending, reducing, changing, or removing an existing logical volume
- Creating a new shared filesystem
- Extending, changing, or removing an existing filesystem
- Adding, removing physical volumes.

When performing any of these maintenance tasks on shared LVM components, make sure that ownership and permissions are reset (on logical volumes) when a volume group is exported and then re-imported. After exporting and importing, a volume group is owned by root and accessible by the system group. Applications, such as some database servers that use raw logical volumes may be affected by this if they change the ownership of the raw logical volume device. You must restore the ownership and permissions to what is needed after this sequence.

Understanding C-SPOC

The C-SPOC commands only operate on both shared and concurrent LVM components that are defined as part of an HACMP resource group. When you use SMIT HACMP C-SPOC, it executes the command on the node that owns the LVM component (the node that has it varied on).

If you run a **ps** command to verify what processes are running during a C-SPOC LVM operation, such as creating, extending, mirroring or unmirroring a shared volume group, you see output similar to the following:

```
ps -ef | grep vg     root 11952 13522   0 08:56:25      -  0:00 ksh
/usr/es/sbin/cluster/cspoc/cexec cllvmcmd -extendvg -f
gdgpgogdhfhchgghdb gigegjhdgldbda.
```

That is a C-SPOC encapsulation of arguments and parameters when data is sent off to remote nodes.

Understanding C-SPOC and Its Relation to Resource Groups

The C-SPOC commands that modify LVM components require a resource group name as an argument. The LVM component that is the target of the command *must* be configured in the resource group specified. C-SPOC uses the resource group information to determine on which nodes it must execute the operation specified.

Removing a Filesystem or Logical Volume

When removing a filesystem or logical volume using C-SPOC, the target filesystem or logical volume must *not* be configured as a resource in the resource group specified. You must remove the configuration for it from the resource group before removing the filesystem or logical volume.

Migrating a Resource Group

You can use the Resource Group Management utility under the **Extended Configuration > System Management (C-SPOC) > HACMP Resource Group and Application Management** menu in SMIT to perform resource group maintenance tasks. This utility enhances failure recovery capabilities of HACMP and allows you to change the status or the location of any type of resource group (along with its resources—IP addresses, applications, and disks), without stopping cluster services. For instance, you can use this utility to free a given node of any resource groups in order to perform system maintenance on that cluster node.

Non-concurrent resource group management tasks that you can perform using the Resource Group Management utility are:

- Dynamically move a specified non-concurrent resource group from a node it currently resides on to the destination node that you have specified.
- Take a non-concurrent resource group online or offline on one or all nodes in the cluster.

For more information on Resource Group Migration, see the section Resource Group Migration in Chapter 15: Managing Resource Groups in a Cluster.

Updating LVM Components in an HACMP Cluster

When you change the definition of a shared LVM component in a cluster, the operation updates the LVM data that describes the component on the local node and in the Volume Group Descriptor Area (VGDA) on the disks in the volume group. AIX 5L LVM enhancements allow all nodes in the cluster to be aware of changes to a volume group, logical volume, and filesystem, at the time the changes are made, rather than waiting for the information to be retrieved during a *lazy update*.

Note: See Lazy Update Processing in an HACMP Cluster for a full explanation of this process.

If for some reason the node is *not* updated via the C-SPOC enhanced utilities, due to an error condition (a node is down, for example), the volume group will be updated and the change will be taken care of during execution of the **clvaryonvg** command.

If node failure does occur during a C-SPOC operation, an error is displayed to the panel and the error messages are recorded in the C-SPOC log. (**/tmp/cspoc.log** is the default location of this log.) Other C-SPOC failures are also logged to the **cspoc.log** but are *not* displayed. You should check this log when any C-SPOC problems occur.

If you change the name of a filesystem, or remove a filesystem and then perform a lazy update, lazy update does *not* run the **imfs -lx** command before running the **imfs** command. This may lead to a failure during fallover or prevent a successful restart of the HACMP cluster services.

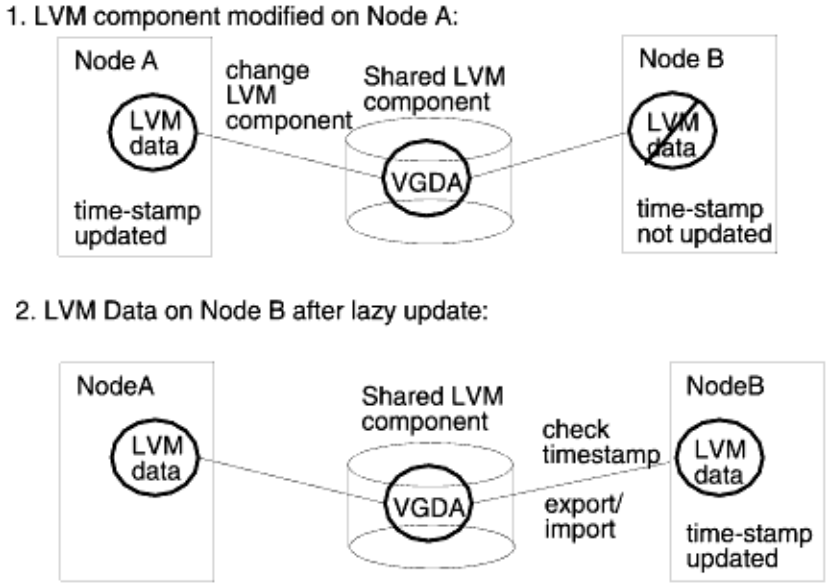
To prevent this from occurring, use the C-SPOC utility to change or remove filesystems. This ensures that **imfs -lx** runs before **imfs** and that the changes are updated on all nodes in the cluster.

Error reporting provides detailed information about inconsistency in volume group state across the cluster. If this happens, you must take manual corrective action. For example, if the filesystem changes are *not* updated on all nodes, update the nodes manually with this information.

Lazy Update Processing in an HACMP Cluster

For LVM components under the control of HACMP, you do *not* have to explicitly do anything to bring the other cluster nodes up to date. Instead, HACMP can perform an **importvg -L** when it activates the volume group during a fallover. (In a cluster, HACMP controls when volume groups are activated.) HACMP implements a function, called *lazy update*, by keeping a copy of the time stamp from the volume group's VGDA. AIX 5L updates this time stamp whenever the LVM component is modified. When another cluster node attempts to vary on the volume group, HACMP compares its copy of the time stamp with the time stamp in the VGDA on the disk. If the values are different, the HACMP software exports and re-imports the volume group before activating it. If the time stamps are the same, HACMP activates the volume group without exporting and re-importing.

The following figure illustrates how a lazy update can be achieved in a cluster using the C-SPOC utility. All nodes in the cluster are updated accordingly, if they are specified from the node originating the command.



Lazy Update Processing

Note: Starting with HACMP 5.2, HACMP does *not* require lazy update processing for enhanced concurrent volume groups, as it keeps all cluster nodes updated with the LVM information.

Forcing an Update before Fallover

In certain circumstances, you may want to update the LVM definition on remote cluster nodes before a fallover occurs. For example, if you rename a logical volume using C-SPOC, the LVM data describing the component is updated on the local node and is updated in the VGDA on the disks, as previously described. If you attempt to rename the logical volume a second time using C-SPOC, the operation fails if the LVM data on any other cluster node has *not* been updated.

When you use C-SPOC to update the LVM data on a remote node, it causes the specified remote node to run **importvg -L** to update the LVM data whether the time stamp associated with the LVM component is the same or different.

Note: The volume group must be varied off on all nodes accessing the shared LVM component before running the update.

Maintaining Shared Volume Groups

While maintaining the HACMP cluster, you may need to perform the following administrative tasks with shared volume groups:

- Enabling Fast Disk Takeover
- Understanding Active and Passive Varyon in Enhanced Concurrent Mode
- Collecting Information on Current Volume Group Configuration
- Importing Shared Volume Groups
- Creating a Shared Volume Group with C-SPOC
- Setting Characteristics of a Shared Volume Group
- Mirroring a Volume Group Using C-SPOC
- Unmirroring a Volume Group Using C-SPOC
- Synchronizing Volume Group Mirrors
- Synchronizing a Shared Volume Group Definition.

Using C-SPOC simplifies the steps required for all tasks. Moreover, you do *not* have to stop and restart cluster services to do the tasks.

Enabling Fast Disk Takeover

HACMP automatically uses fast disk takeover for enhanced concurrent mode volume groups that are included as resources in shared resource groups residing on shared disks.

If you upgraded from previous releases and have existing volume groups included in non-concurrent resource groups, to use this functionality, convert these volume groups to enhanced concurrent volume groups using C-SPOC.

Fast disk takeover is supported when AIX 5L v.5.2 and greater is installed on all nodes in the cluster. If you include your enhanced concurrent volume groups in the shared resource groups *and* AIX 5L v.5.2 or greater is *not* detected as being installed on all nodes in the cluster, you receive an error upon cluster verification.

To use fast disk takeover after an upgrade from a previous release, use C-SPOC to convert existing volume groups included in non-concurrent resource groups to enhanced concurrent volume groups.

For more information on fast disk takeover, see Chapter 4: Planning Shared LVM Components in the *Planning Guide.*

Understanding Active and Passive Varyon in Enhanced Concurrent Mode

An enhanced concurrent volume group can be made active on the node, or varied on, in two states: active or passive. Note that active or passive state varyons are done automatically by HACMP upon detection of the enhanced concurrent mode volume group, based on the state of the volume group and current cluster configuration.

WARNING: All nodes in the cluster must be available before making any LVM changes. This ensures that all nodes have an accurate view of the state of the volume group. For more information about safely performing a forced varyon operation, and on instructions how to configure it in SMIT, see Forcing a Varyon of Volume Groups in Chapter 5: Configuring HACMP Resource Groups (Extended).

Active State Varyon

Active state varyon behaves as ordinary varyon, and makes the logical volumes normally available. When an enhanced concurrent volume group is varied on in active state on a node, it allows the following operations:

- Operations on filesystems, such as filesystem mounts
- Operations on applications
- Operations on logical volumes, such as creating logical volumes
- Synchronizing volume groups.

Passive State Varyon

When an enhanced concurrent volume group is varied on in passive state, the LVM provides an equivalent of fencing for the volume group at the LVM level.

Passive state varyon allows only a limited number of read-only operations on the volume group:

- LVM read-only access to the volume group's special file
- LVM read-only access to the first 4k of all logical volumes that are owned by the volume group.

The following operations are *not* allowed when a volume group is varied on in passive state:

- Operations on filesystems, such as filesystems mounting
- Any operations on logical volumes, such as having logical volumes open
- Synchronizing volume groups.

Using Active or Passive State Varyon in HACMP

HACMP detects when a volume group included in a shared resource group is converted to or defined as an enhanced concurrent mode volume group, and notifies the LVM which node currently owns the volume group. Based on this information, the LVM activates the volume group in the appropriate active or passive state depending on the node on which this operation takes place.

- Upon cluster startup, if the volume group resides currently on the node that owns the resource group, HACMP activates the volume group on this node in active state. HACMP activates the volume group in passive state on all other nodes in the cluster. Note that HACMP will activate a volume group in active state only on one node at a time.
- Upon fallover, if a node releases a resource group, or, if the resource group is being moved to another node for any other reason, HACMP switches the varyon state for the volume group from active to passive on the node that releases the resource group (if cluster services are still running), and activates the volume group in active state on the node that acquires the resource group. The volume group remains in passive state on all other nodes in the cluster.
- Upon node reintegration, this procedure is repeated. HACMP changes the varyon state of the volume group from active to passive on the node that releases the resource group and varies on the volume group in active state on the joining node. While activating, the volume group remains passive on all other nodes in the cluster.

Note: The switch between active and passive states is necessary to prevent mounting filesystems on more than one node at a time.

See Chapter 12: Managing Shared LVM Components in a Concurrent Access Environment for information on enhanced concurrent volume groups in concurrent resource groups.

Verification Checks for Shared Volume Groups Defined for Auto Varyon

Typically, shared volume groups listed within a resource group should have their **auto-varyon** attribute in the AIX 5L ODM set to **No**.

The HACMP verification checks that the volume group **auto-varyon** flag is set to **No**. If you use the interactive mode for verification, you will be prompted to set the **auto-varyon** flag to **No** on all the cluster nodes listed in the resource group.

Checking the Status of a Volume Group

As with regular cluster takeover operations, you can debug and trace the cluster activity for fast disk takeover using the information logged in the **hacmp.out** file. You may check the status of the volume group by issuing the **lsvg** command. Depending on your configuration, the **lsvg** command returns the following settings:

- `VG STATE` will be `active` if it is varied on either actively or passively.
- `VG PERMISSION` will be `read/write` if it is actively varied on on the node, or `passive-only`, if it is passively varied on.
- `CONCURRENT` will either be `Capable` or `Enhanced-Capable (for concurrent volume groups).`

Here is an example of **lsvg** output:

```
# lsvg vg1

VOLUME GROUP:   vg1                     VG IDENTIFIER:
00020adf00004c00000000f329382713
VG STATE:       active                  PP SIZE:        16 megabyte(s)
VG PERMISSION:  passive-only            TOTAL PPs:      542 (8672
megabytes)
MAX LVs:        256                     FREE PPs:       521 (8336
megabytes)
```

```
LVs:            3                   USED PPs:       21 (336
megabytes)
OPEN LVs:       0                   QUORUM:         2
TOTAL PVs:      1                   VG DESCRIPTORS: 2
STALE PVs:      0                   STALE PPs:      0
ACTIVE PVs:     1                   AUTO ON:        no
Concurrent:     Enhanced-Capable    Auto-Concurrent: Disabled
VG Mode:        Concurrent
Node ID:        2                   Active Nodes:   1 4
MAX PPs per PV: 1016                MAX PVs:        32
LTG size:       128 kilobyte(s)     AUTO SYNC:      no
HOT SPARE:      no                  BB POLICY:      relocatable
```

Avoiding a Partitioned Cluster

When configuring enhanced concurrent volume groups in shared resource groups, ensure that multiple networks exist for communication between the nodes in the cluster, to avoid cluster partitioning. When fast disk takeover is used, the normal SCSI reserve is *not* set to prevent multiple nodes from accessing the volume group.

In a partitioned cluster, it is possible that nodes in each partition could accidentally vary on the volume group in active state. Because active state varyon of the volume group allows filesystem mounts and changes to the physical volumes, this state can potentially lead to different copies of the same volume group. Make sure that you configure multiple communication paths between the nodes in the cluster.

Collecting Information on Current Volume Group Configuration

HACMP can collect information about all shared volume groups that are currently available on the nodes in the cluster, *and* volume groups that can be imported to the other nodes in the resource group. HACMP filters out volume groups that are already included in any of the resource groups.

You can use this information to import discovered volume groups onto other nodes in the resource group that do *not* have these volume groups.

To collect the information about volume group configuration:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Discover HACMP-related Information from Configured Nodes** and press Enter.

   The information on current volume group configuration is collected and displayed.

Importing Shared Volume Groups

When adding a volume group to the resource group, you may choose to manually import a volume group onto the destination nodes, or you may automatically import it onto all the destination nodes in the resource group.

Importing a Volume Group Automatically

You can set automatic import of a volume group in SMIT under the **HACMP Extended Resource Group Configuration** menu. It enables HACMP to automatically import shareable volume groups onto all the destination nodes in the resource group. Automatic import allows you to create a volume group and then add it to the resource group immediately, without manually importing it onto each of the destination nodes in the resource group.

Prior to importing volume groups automatically, make sure that you have collected the information on appropriate volume groups, using the **HACMP Extended Configuration > Discover HACMP-related Information from Configured Nodes** action in SMIT.

Note: Each volume group is assigned a major number when it is created. When HACMP automatically imports a volume group, the major number already assigned to the volume group will be used if it is available on all the destination nodes. Otherwise, any free major number will be used.

Prerequisites and Notes

In order for HACMP to import available volume groups, ensure the following conditions are met:

- Volume group names must be the same across cluster nodes and unique to the cluster.
- Logical volumes and filesystems must have unique names.
- All physical disks must be known to AIX 5L and have PVIDs assigned.
- The physical disks on which the volume group resides are available to all of the nodes in the resource group.

Procedure for Importing a Volume Group Automatically

To add volume groups to a resource group via automatic import:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Resource Configuration > HACMP Extended Resource Group Configuration > Change/Show Resources/Attributes for a Resource Group** and press Enter.
3. On the next panel, select the resource group for which you want to define a volume group and press Enter. A panel appears with fields for all the types of resources applicable for the type of resource group you selected.
4. In the **Volume Groups** field, you can select the volume groups from the picklist or enter volume groups names.

   If, prior to this procedure, you requested that HACMP collect information about the appropriate volume groups, then pressing the F4 key gives you a list of all volume groups collected cluster-wide, including all shared volume groups in the resource group *and* the volume groups that are currently available for import onto the resource group nodes. HACMP filters out from this list those volume groups that are already included in any of the resource groups.

Note: The list of volume groups will include only the non-concurrent capable volume groups. This list will *not* include **rootvg** and any volume groups already defined to another resource group.

5. Set the **Automatically Import Volume Groups** flag to **True**. (The default is **False**.)
6. Press Enter. HACMP determines whether the volume groups that you entered or selected in the **Volume Groups** field need to be imported to any of the nodes that you defined for the resource group and proceeds to import them as needed.

Final State of Automatically Imported Volume Groups

When HACMP automatically imports volume groups, their final state (varied on or varied off) depends on the initial state of the volume group and whether the resource group is online or offline when the import occurs.

In all cases, the volume group ends up varied on after the resource group is started or after the cluster resources are synchronized, even if it is varied off at some point during the import process.

This table shows the initial condition of the volume group after its creation, the state of the resource group when the import occurs, and the resulting state of the imported volume group:

Initial Volume Group State	Resource Group State	Auto Imported Volume Group State
Varied ON	OFFLINE	Remains varied ON
Varied ON	ONLINE	Varied ON
Varied OFF	OFFLINE	Varied OFF until the resource group is started
Varied OFF	ONLINE	Varied ON

Importing a Volume Group Manually

If you want to import your volume group manually upon adding it to the resource group, make sure that in SMIT the **Automatically Import Volume Groups** flag is set to **False** (this is the default) and use the AIX 5L **importvg** fastpath.

Importing a Shared Volume Group with C-SPOC

To import a volume group using the C-SPOC utility:

1. Complete prerequisite tasks. The physical volumes (**hdisks**) in the volume group must be installed, configured, and available on all nodes that can own the volume group.
2. On any cluster node that can own the shared volume group (is in the participating nodes list for the resource group), vary on the volume group, using the SMIT **varyonvg** fastpath (if it is *not* varied on already).
3. On the source node, enter the fastpath `smit cl_admin`
4. In SMIT, select **HACMP Logical Volume Management > Shared Volume Groups > Import a Shared Volume Group** and press Enter.

   A list of volume groups appears. (Enhanced concurrent volume groups are also included as choices in picklists for non-concurrent resource groups.)

5. Select a volume group and press Enter.

   A list of physical volumes appears.

6. Select a physical volume and Press Enter.

   SMIT displays the **Import a Shared Volume Group** panel. Values for fields you have selected are displayed.

7. Enter values for other fields as follows:

Resource Group name	The cluster resource group to which this shared volume group belongs.
VOLUME GROUP name	The name of the volume group that you are importing.
PHYSICAL VOLUME name	The name of one of the physical volumes that resides in the volume group. This is the hdisk name on the reference node.
Reference node	The node from which the physical disk was retrieved.
Volume Group MAJOR NUMBER	If you are *not* using NFS, use the default (which is the next available number in the valid range). If you are using NFS, you must be sure to use the same major number on all nodes. Use the **lvlstmajor** command on each node to determine a free major number common to all nodes.
Make this VG concurrent capable?	For a non-concurrent volume group, set this field to **no**. The default is **no**.
Make default varyon of VG Concurrent?	For a non-concurrent volume group, set this field to **no**. The default is **no**.

8. If this panel reflects the correct information, press Enter to import the shared volume group. All nodes in the cluster receive this updated information.

   If you did this task from a cluster node that does *not* need the shared volume group varied on, vary off the volume group on that node.

Creating a Shared Volume Group with C-SPOC

Before creating a shared volume group for the cluster using C-SPOC, check that:

- All disk devices are properly attached to the cluster nodes.
- All disk devices are properly configured on all cluster nodes and the devices are listed as available on all nodes.
- Disks have a PVID.
- If you create concurrent SSA volume groups, you must assign unique non-zero node numbers with `ssar` on each cluster node. If you select the option for SSA fencing, HACMP automatically assigns unique node numbers during cluster synchronization.

Note: If you add a VPATH disk to a volume group made up of hdisks, the volume group will be converted to VPATHs on all nodes.

To create a shared volume group for a selected list of cluster nodes:

1. Enter the fastpath `smit cl_admin`
2. In SMIT, select **HACMP Logical Volume Management** > **Shared Volume Groups** > **Create a Shared Volume Group**. (Or **Create a Shared Volume Group with Data Path Devices**) and press Enter.

   SMIT displays a list of cluster nodes.
3. Select two or more nodes from the list and press Enter.

   The system correlates a list of all free physical disks that are available to all nodes selected. (Free disks are those disks that currently are *not* part of a volume group and have PVIDs.) SMIT displays the list of free physical disks in a multi-picklist by PVIDs. If you are creating a shared volume group on Data Path Devices, only VPATH-capable disks are listed.
4. Select one or more disks from the list and press Enter.

   SMIT displays the **Add a Volume Group** panel.
5. Complete the selections as follows and press Enter.

Node Names	The node(s) you selected.
PVID	PVID of the selected disk.
VOLUME GROUP name	The name of the volume group must be unique within the cluster and distinct from the service IP label/address and resource group names; it should relate to the application it serves, as well as to any corresponding device. For example, `websphere_service_VG`.
Physical partition SIZE in megabytes	Accept the default.
Volume group MAJOR NUMBER	The system displays the number that C-SPOC has determined to be correct. **WARNING**: Changing the volume group major number may result in the command's inability to execute on a node that does *not* have that major number currently available. Please check for a commonly available major number on all nodes before changing this setting.
Enable Cross-Site LVM Mirroring	Set this field to **True** to enable data mirroring between sites. The default is **False**. See Configuring Cross-Site LVM Mirroring for more information.

C-SPOC verifies communication paths and version compatibility and then executes the command on all nodes in selection. If cross-site LVM mirroring is enabled, that configuration is verified.

Note: If the major number that you entered on the SMIT panel was *not* free at the time that the system attempted to make the volume group, HACMP displays an error for the node that did *not* complete the command, and continues to the other nodes. At the completion of the command the volume group will *not* be active on any node in the cluster.

6. Run the discovery process so that the new volume group is included in picklists for future actions.

Setting Characteristics of a Shared Volume Group

You can change the volume group's characteristics by:

- Adding or removing a volume from the shared volume group
- Enabling or disabling the volume group for cross-site LVM mirroring.

Adding or Removing a Volume from a Shared Volume Group

To add or remove a volume to or from a shared volume group:

1. Enter the fastpath `smit cl_admin`
2. In SMIT, select **HACMP Logical Volume Management** > **Shared Volume Groups** > **Set Characteristics of a Shared Volume Group** > **Add (or Remove) a Volume from a Shared Volume Group** and press Enter.

   SMIT displays a list of volume groups.
3. Select the volume group and press Enter.
4. Select the volume to add or remove from the list and press Enter.
5. Synchronize the cluster.

Enabling or Disabling Cross-Site LVM Mirroring

For more information, see Configuring Cross-Site LVM Mirroring.

To enable or disable cross-site LVM mirroring of a shared volume group:

1. Enter the fastpath `smit cl_admin`
2. In SMIT, select **HACMP Logical Volume Management** > **Shared Volume Groups** > **Set Characteristics of a Shared Volume Group > Enable/Disable a Shared Volume Group for Cross-site Mirroring** and press Enter.

   SMIT displays a list of volume groups.
3. Select a volume group from the list.
4. Enable or disable the field **Enable Cross-Site LVM Mirroring** (**True** or **False**) and press Enter. If you disable cross-site mirroring, do this on each node that can own the volume group.
5. Synchronize the cluster if the cluster services are *not* running. If you have sites defined, you cannot synchronize while cluster services are running.

Mirroring a Volume Group Using C-SPOC

To mirror a shared volume group using the C-SPOC utility:

1. Complete prerequisite tasks. The physical volumes (**hdisks**) in the volume group must be installed, configured, and available.
2. On any cluster node that can own the shared volume group (is in the participating nodes list for the resource group), vary on the volume group, using the SMIT **varyonvg** fastpath (if it is *not* varied on already).
3. On the source node, enter `smit cl_admin`
4. In SMIT, select **HACMP Logical Volume Management > Shared Volume Groups > Mirror a Shared Volume Group**.

   SMIT displays a list of resource groups and associated volume groups. (Enhanced concurrent volume groups are also included as choices in picklists for non-concurrent resource groups.)
5. Select a volume group and press Enter.
6. Select entries from the list of nodes and physical volumes (hdisks) and Press Enter.

   SMIT displays the **Mirror a Shared Volume Group** panel, with the selected entries filled in.
7. Enter values for other fields as follows:

Resource Group Name	SMIT displays the name of the resource group to which this shared volume group belongs.
VOLUME GROUP name	SMIT displays the name of the volume group that you selected to mirror.
Reference node	SMIT displays the node from which the name of the physical disk was retrieved.
VOLUME names	The name of a physical volume on the volume group that you selected to mirror. This is the **hdisk** name on the reference node.
FORCE deallocation of all partitions on this physical volume?	The default is **no**.
Number of COPIES of each logical partition	Select **2** or **3**. The default is **2**.
Keep Quorum Checking On?	You can also select **yes** or **no**.The default is **no.**
Create Exact LV Mapping?	The default is **no**.

8. If this panel reflects the correct information, press Enter to mirror the shared volume group. All nodes in the cluster receive this updated information.

If you did this task from a cluster node that does *not* need the shared volume group varied on, vary off the volume group on that node.

Unmirroring a Volume Group Using C-SPOC

To unmirror a shared volume group using the C-SPOC utility:

1. Complete prerequisite tasks. The physical volumes (**hdisks**) in the volume group must be installed, configured, and available.
2. On any cluster node that can own the shared volume group (is in the participating nodes list for the resource group), vary on the volume group, using the SMIT **varyonvg** fastpath (if it is *not* varied on already).
3. On the source node, enter the fastpath `smit cl_admin`
4. In SMIT, select **HACMP Logical Volume Management > Shared Volume Groups > Unmirror a Shared Volume Group** and press Enter.

   SMIT displays a list of resource groups and associated volume groups. (Enhanced concurrent volume groups are also included as choices in picklists for non-concurrent resource groups.)
5. Select a volume group and press Enter.
6. Select entries from the list of nodes and physical volumes (**hdisks**) and Press Enter.

   SMIT displays the Unmirror a Shared Volume Group panel, with the chosen fields filled in.
7. Enter values for other fields as follows:

Resource Group Name	SMIT displays the name of the resource group to which this shared volume group belongs.
VOLUME GROUP name	SMIT displays the name of the volume group that you selected to unmirror.
Reference node	SMIT displays the node from which the name of the physical disk was retrieved.
VOLUME names	SMIT displays the name of a physical volume on the volume group that you selected to unmirror. This is the hdisk name on the reference node.
Mirror sync mode	Select **Foreground**, **Background**, or **No Sync**. **Foreground** is the default.
Number of COPIES of each logical partition	Select **2** or **3**. The default is **2**.

8. If this panel reflects the correct information, press Enter to unmirror the shared volume group. All nodes in the cluster receive this updated information.

   If you did this task from a cluster node that does *not* need the shared volume group varied on, vary off the volume group on that node.

Synchronizing Volume Group Mirrors

To synchronize shared LVM Mirrors by volume group using the C-SPOC utility:

1. Complete prerequisite tasks. The physical volumes (**hdisks**) in the volume group must be installed, configured, and available.
2. On any cluster node that can own the shared volume group (is in the participating nodes list for the resource group), vary on the volume group, using the SMIT **varyonvg** fastpath (if it is *not* varied on already).
3. On the source node, enter the fastpath `smit cl_admin`
4. In SMIT, select **HACMP Logical Volume Management > Synchronize Shared LVM Mirrors > Synchronize By Volume Group** and press Enter.

   SMIT displays a list of resource groups and associated volume groups. (Enhanced concurrent volume groups are also included as choices in picklists for non-concurrent resource groups.)
5. Select a volume group and press Enter.

   SMIT displays the **Synchronize LVM Mirrors by Volume Group** panel, with the chosen entries filled in.
6. Enter values for other fields as follows:

Resource Group Name	SMIT displays the name of the resource group to which this shared volume group belongs.
VOLUME GROUP name	SMIT displays the name of the volume group that you selected to mirror.
Reference node	SMIT displays the node from which the name of the physical disk was retrieved.
Number of Partitions to Sync in Parallel	Leave empty.
Synchronize All Partitions	The default is **no.**
Delay Writes to VG from other cluster nodes during this Sync	The default is **no**.

7. If this panel reflects the correct information, press Enter to synchronize LVM mirrors by the shared volume group. All nodes in the cluster receive this updated information.

   If you did this task from a cluster node that does *not* need the shared volume group varied on, vary off the volume group on that node.

Synchronizing a Shared Volume Group Definition

To synchronize a shared volume group definition using the C-SPOC utility:

1. Complete prerequisite tasks. The physical volumes (**hdisks**) in the volume group must be installed, configured, and available.
2. On the source node, enter the fastpath `smit cl_admin`

3. In SMIT, select **HACMP Logical Volume Management > Synchronize a Shared Volume Group Definition** and press Enter.
4. On the **Synchronize a Shared Volume Group Definition** panel, enter the name of the shared volume group to synchronize. Press F4 for a picklist. The picklist contains only volume groups that are *not* varied on anywhere in the cluster.
5. Select a volume group and press Enter.

   The command runs. All nodes in the cluster receive this updated information.

Maintaining Logical Volumes

The following administrative tasks involve shared logical volumes. You can perform all these tasks using the C-SPOC utility:

- Adding a Logical Volume to a Cluster Using C-SPOC
- Setting Characteristics of a Shared Logical Volume Using C-SPOC
- Changing a Shared Logical Volume
- Removing a Logical Volume Using C-SPOC
- Synchronizing LVM Mirrors by Logical Volume.

Note: On RAID devices, increasing or decreasing the number of copies (mirrors) of a shared logical volume is *not* supported.

Adding a Logical Volume to a Cluster Using C-SPOC

To add a logical volume to a cluster using C-SPOC:

1. Enter the C-SPOC fastpath: `smit cl_admin`
2. In SMIT, select **Cluster Logical Volume Manager** > **Shared Logical Volumes** > **Add a Shared Logical Volume** and press Enter.

   SMIT displays a list of resource groups and associated shared volume groups.
3. Select a resource group-volume group combination and press Enter.

   SMIT displays a list of physical volumes.
4. Select a physical volume and press Enter. The **Add a Shared Logical Volume** panel appears, with chosen fields filled in as shown in the sample below:

Resource Group name	casc_rg
VOLUME GROUP name	sharedvg
Reference node	a1
Number of LOGICAL PARTITIONS	[]
PHYSICAL VOLUME names	hdisk16

Logical volume NAME	If you are configuring the logical volume for cross-site LVM mirrors, include an appropriate name, for example, *lv1site1*.
Logical volume TYPE	[]
POSITION on physical volume	middle
RANGE of physical volumes	minimum
MAXIMUM NUMBER of PHYSICAL VOLUMES to use for allocation	[]
Number of COPIES of each logical partition	Include at least 2 for cross-site LVM mirror configurations.
Mirror Write Consistency?	no
Allocate each logical partition copy on a SEPARATE physical volume?	**yes** is the default. If you specify a forced varyon attribute in SMIT for volume groups in a resource group, it is recommended to set this field to **super strict**.
RELOCATE the logical volume during reorganization	yes
Logical volume LABEL	[]
MAXIMUM NUMBER of LOGICAL PARTITIONS	[512]
Enable BAD BLOCK relocation?	no
SCHEDULING POLICY for reading/writing logical partition copies	parallel
Enable WRITE VERIFY?	no
File containing ALLOCATION MAP	[]
Stripe Size?	[Not Striped]

5. The default logical volume characteristics are most common. Make changes if necessary for your system and press Enter. Other cluster nodes are updated with this information.

Setting Characteristics of a Shared Logical Volume Using C-SPOC

This section contains instructions for the following tasks that you can do for all cluster nodes from one node with C-SPOC:

- Renaming a Shared Logical Volume Using C-SPOC

- Increasing the Size of a Shared Logical Volume Using C-SPOC
- Adding a Copy to a Shared Logical Volume Using C-SPOC
- Removing a Copy from a Shared Logical Volume Using C-SPOC.

Renaming a Shared Logical Volume Using C-SPOC

To rename a shared logical volume on all nodes in a cluster by executing a C-SPOC command on any node:

1. Enter the fastpath `smit cl_admin`
2. In SMIT, select **HACMP Logical Volume Management** > **Shared Logical Volumes > Set Characteristics of A Shared Logical Volume > Rename a Shared Logical Volume** and press Enter. SMIT displays the **Rename a Logical Volume on the Cluster** panel and press Enter.

   SMIT displays a picklist of existing logical volumes.
3. Select a logical volume and press Enter. SMIT displays a panel with the **Resource group name** and **Current logical volume name** filled in.
4. Enter the new name in the **NEW logical volume name** field and press Enter. The C-SPOC utility changes the name on all cluster nodes.

Note: After completing this procedure, *confirm your changes* by initiating failures and verifying correct fallover behavior before resuming normal cluster operations.

Increasing the Size of a Shared Logical Volume Using C-SPOC

To increase the size of a shared logical volume on all nodes in a cluster:

1. On any node, enter the SMIT fastpath: `smit cl_admin`
2. In SMIT, select **HACMP Logical Volume Management** > **Shared Logical Volumes > Set Characteristics of A Shared Logical Volume > Increase the Size of a Shared Logical Volume** and press Enter. SMIT displays a list of logical volumes arranged by resource group.
3. Select a logical volume from the picklist and press Enter.

   SMIT displays a list of physical volumes.
4. Select a physical volume and press Enter. SMIT displays the **Increase Size of a Shared Logical Volume** panel with the **Resource Group**, **Logical Volume**, **Reference Node** and default fields filled.
5. Enter the new size in the **Number of ADDITIONAL logical partitions** field and press Enter. The C-SPOC utility changes the size of this logical volume on all cluster nodes.

Adding a Copy to a Shared Logical Volume Using C-SPOC

To add a copy to a shared logical volume on all nodes in a cluster:

1. On any node, enter the fastpath `smit cl_admin`

2. In SMIT, select **HACMP Logical Volume Management** > **Shared Logical Volumes > Set Characteristics of A Shared Logical Volume > Add a Copy to a Shared Logical Volume** and press Enter. SMIT displays a list of logical volumes arranged by resource group.
3. Select a logical volume from the picklist and press Enter. SMIT displays a list of physical volumes.
4. Select a physical volume and press Enter. SMIT displays the **Add a Copy to a Shared Logical Volume** panel with the **Resource Group**, **Logical Volume**, **Reference Node** and default fields filled.
5. Enter the new number of mirrors in the **NEW TOTAL number of logical partitions** field and press Enter. The C-SPOC utility changes the number of copies of this logical volume on all cluster nodes.

Removing a Copy from a Shared Logical Volume Using C-SPOC

To remove a copy of a shared logical volume on all nodes in a cluster:

1. On any node, enter the fastpath `smit cl_admin`
2. In SMIT, select **HACMP Logical Volume Management** > **Shared Logical Volumes > Set Characteristics of A Shared Logical Volume > Remove a Copy from a Shared Logical Volume** and press Enter. SMIT displays a list of logical volumes arranged by resource group.
3. Select the logical volume from the picklist and press Enter. SMIT displays a list of nodes and physical volumes.
4. Select the physical volumes from which you want to remove copies and press Enter. SMIT displays the **Remove a Copy from a Shared Logical Volume** panel with the **Resource Group, Logical Volume name, Reference Node** and **Physical Volume** names fields filled in.
5. Enter the new number of mirrors in the **NEW maximum number of logical partitions copies** field and check the **PHYSICAL VOLUME name(s) to remove copies from** field to make sure it is correct and press Enter. The C-SPOC utility changes the number of copies of this logical volume on all cluster nodes.
6. To check the status of the C-SPOC command execution on all nodes, view the C-SPOC log file in **/tmp/cspoc.log**.

Changing a Shared Logical Volume

To change the characteristics of a shared logical volume on all nodes in a cluster:

1. On any node, enter the fastpath `smit cl_admin`
2. In SMIT, select HACMP **Logical Volume Management** > **Shared Logical Volumes > Change a Shared Logical Volume** option and press Enter. SMIT displays a picklist of existing logical volumes.
3. Select the logical volume. SMIT displays the panel, with the values of the selected logical volume attributes filled in.

4. Enter data in the fields you want to change and press Enter. The C-SPOC utility changes the characteristics on the local node. The logical volume definition is updated on remote nodes.
5. To check the status of the C-SPOC command execution on all nodes, view the C-SPOC log file in **/tmp/cspoc.log**.

Note: After completing this procedure, *confirm your changes* by initiating failures and verifying correct fallover behavior before resuming normal cluster operations.

Removing a Logical Volume Using C-SPOC

Note: If the logical volume to be removed contains a filesystem, you first must remove the filesystem from any specified resource group before attempting to remove the logical volume. Afterwards, be sure to synchronize cluster resources on all cluster nodes.

To remove a logical volume on any node in a cluster:

1. On any node, enter the fastpath `smit cl_admin`
2. In SMIT, select **HACMP Logical Volume Management > Shared Logical Volumes** > **Remove a Shared Logical Volume** and press Enter.
3. C-SPOC provides a list of shared logical volumes, organized by HACMP resource group. Select the logical volume you want to remove and press Enter. Remote nodes are updated.
4. To check the status of the C-SPOC command execution on other cluster nodes, view the C-SPOC log file in **/tmp/cspoc.log**.

Synchronizing LVM Mirrors by Logical Volume

To synchronize shared LVM mirrors by logical volume using the C-SPOC utility:

1. Complete prerequisite tasks. The physical volumes (**hdisks**) in the volume group must be installed, configured, and available.
2. On any cluster node that can own the shared volume group (is included in the participating nodes list for the resource group), vary on the volume group, using the SMIT **varyonvg** fastpath (if it is *not* varied on already).
3. On the source node, enter the fastpath `smit cl_admin`
4. In SMIT, select **HACMP Logical Volume Management > Synchronize Shared LVM Mirrors > Synchronize By Logical Volume** and press Enter.

   SMIT displays a list of resource groups and associated logical volumes.
5. Select a logical volume and press Enter.

   SMIT displays the **Synchronize LVM Mirrors by Volume Group** panel, with the chosen entries filled in.

6. Enter values for other fields as follows:

Resource Group Name	SMIT displays the name of the resource group to which this logical volume belongs.
LOGICAL VOLUME name	SMIT displays the name of the logical volume that you selected to synchronize.
Number of Partitions to Sync in Parallel	Leave empty.
Synchronize All Partitions	The default is **no.**
Delay Writes to VG from other cluster nodes during this Sync	The default is **no**.

7. If this panel reflects the correct information, press Enter to synchronize LVM mirrors by the shared logical volume. All nodes in the cluster receive this updated information.
8. If you did this task from a cluster node that does *not* need the shared volume group varied on, vary off the volume group on that node.

Maintaining Shared Filesystems

The following administrative tasks involve shared filesystems:

- Creating Shared Filesystems with C-SPOC
- Adding the Filesystem to an HACMP Cluster Logical Volume
- Changing a Shared Filesystem in HACMP Using C-SPOC
- Removing a Shared Filesystem Using C-SPOC.

Each operation is described below. The sections also describe how to use the C-SPOC utility to create, change or remove a shared filesystem in a cluster.

Journaled Filesystem and Enhanced Journaled Filesystem

Enhanced Journaled Filesystem (JFS2) provides the capability to store much larger files than the Journaled File System (JFS). Additionally, it is the default filesystem for the 64-bit kernel. You can choose to implement either JFS, which is the recommended filesystem for 32-bit environments, or JFS2, which offers 64-bit functionality.

Note: Unlike the JFS filesystem, the JFS2 filesystem will *not* allow the **link()** API to be used on files of type **directory.** This limitation may cause some applications that operate correctly on a JFS filesystem to fail on a JFS2 filesystem.

See the AIX 5L documentation for more information.

The SMIT paths shown in the following sections of this chapter use the Journaled Filesystem; similar paths exist for the Enhanced Journaled Filesystem.

Reliable NFS Server and Enhanced Journaled Filesystem

You can use either JFS or JFS2 filesystems with the Reliable NFS Server functionality of HACMP. JFS2 is supported with a highly available NFS Server functionality only on AIX 5L 5.2 with one of the following minimum fileset levels installed:

```
bos.up.5.2.0.10
bos.mp.5.2.0.10
bos.mp64.5.2.0.10.
```

Without these minimum fileset levels specified, you can specify JFS2 filesystems as HACMP NFS exported filesystems, but the highly available functionality of saving the dup cache for these filesystems is *not* supported. (The lock information transfer functionality is supported.)

Also, note that Reliable NFS Server functionality is supported only with JFS2 filesystems with external logs (jfs2logs). Filesystems with embedded logs are *not* supported.

Creating Shared Filesystems with C-SPOC

Before creating a journaled filesystem for the cluster using C-SPOC, check that:

- All disk devices are properly attached to the cluster nodes
- All disk devices are properly configured and available on all cluster nodes
- The volume group that will contain the filesystem must be varied on, on at least one cluster node.

You can add a journaled filesystem to either:

- A shared volume group (no previously defined cluster logical volume)
- A previously defined cluster logical volume (on a shared volume group).

To add a filesystem where no logical volume is currently defined:

1. Enter the fastpath `smitty cl_admin`
2. In SMIT, select **HACMP Logical Volume Management** > **Shared Filesystems** > **Add a Journaled Filesystem** and press Enter.

   SMIT displays a list of filesystem types (**Standard**, **Compressed** or **Large File Enabled**).
3. Select a filesystem type from the list.

   SMIT generates a list of all volume groups and associated nodes in the cluster.
4. Select the volume group where the filesystem will be added. SMIT displays the AIX 5L SMIT panel for selecting filesystem attributes.
5. Enter field values as follows:

Node Names	SMIT displays the names of the selected cluster nodes.
Volume Group Name	SMIT displays the selected volume group name.
SIZE of filesystem	Set as needed
MOUNT POINT	Enter the mount point for the filesystem.
PERMISSIONS	Set as needed.

Mount OPTIONS	Set as needed.
Start Disk Accounting?	Set as needed. Default is **no**.
Fragment Size (Bytes)	4096 is the default.
Number of Bytes per inode	4096 is the default.
Allocation Group Size (MBytes)	8 is the default.

6. Select the filesystem attributes and press Enter.

   SMIT checks the nodelist for the resource group that contains the volume group, creates the logical volume (on an existing log logical volume if present, otherwise it will create a new log logical volume) and adds the filesystem to the node where the volume group is varied on. All other nodes in the resource group will run an **importvg -L**.

Adding the Filesystem to an HACMP Cluster Logical Volume

To add a filesystem to a previously defined cluster logical volume:

1. Enter the fastpath `smitty cl_admin`
2. In SMIT, select **HACMP Logical Volume Management** > **Shared Filesystems** > **Add a Journaled Filesystem to a Previously Defined Logical Volume** and press Enter. SMIT displays a list of filesystem types (**Standard**, **Compressed**, or **Large File Enabled**).
3. Select the filesystem type from the list. SMIT generates a list of all free logical volumes in the cluster and nodes they are on. SMIT reports a logical volume as free if:
   - The logical volume is part of a parent volume group that is configured as a resource in the cluster
   - The logical volume is varied on prior to and during the system polling the disk for logical volume information
   - The logical volume does *not* have a filesystem mount point.
4. Select a logical volume where the filesystems will be added. SMIT displays the AIX 5L SMIT panel for selecting filesystem attributes.
5. Enter field values as follows:

Node Names	SMIT displays the names of the selected cluster nodes.
LOGICAL VOLUME name	SMIT displays the name of the selected logical volume.
***MOUNT POINT**	Enter the mount point for the filesystem.
PERMISSIONS	Set as needed.
Mount OPTIONS	Set as needed.
Start Disk Accounting?	Set as needed. Default is **no**.
Fragment Size (Bytes)	4096 is the default.

Number of Bytes per inode	4096 is the default.
Allocation Group Size (MBytes)	8 is the default.

6. Select the filesystem attributes and press Enter. SMIT checks the nodelist for the resource group that contains the volume group where the logical volume is located and adds the filesystem to the node where the volume group is varied on. All other nodes in the resource group will run an **importvg -L**.

Changing a Shared Filesystem in HACMP Using C-SPOC

As system administrator of an HACMP cluster, you may need to change the characteristics of an existing filesystem. Using the C-SPOC utility, you can change the characteristics of a shared filesystem on cluster nodes by executing a command on a single cluster node. The C-SPOC command changes the attributes of the shared filesystem on all the nodes in the resource group.

To change the characteristics of a shared filesystem:

1. Vary on the volume group, if needed, by using the **varyonvg** command. You can use the C-SPOC utility to change a filesystem even if the volume group on which it is defined is varied off.
2. Enter the SMIT fastpath `smit cl_admin`
3. In SMIT, select the **HACMP Logical Volume Management** > **Shared Filesystems > Change/Show Characteristics of a Shared Filesystem in the Cluster** and press Enter.

   SMIT displays a picklist of existing filesystems.
4. Select the filesystem to change.

   SMIT displays a panel containing the characteristics of the filesystem.
5. Enter data in the fields to change and press Enter. The C-SPOC utility changes the filesystem characteristics on all nodes in the resource group.
6. (*Optional*) To check the status of the C-SPOC command execution on cluster nodes, view the C-SPOC log file in **/tmp/cspoc.log**.

Removing a Shared Filesystem Using C-SPOC

As system administrator of an HACMP cluster, you may need to remove a filesystem. You can optionally remove the filesystem's mount point as part of the same operation. Using the following procedure, you can remove a shared filesystem on any node in a cluster.

C-SPOC deletes the shared filesystem on the node that currently has the shared volume group varied on. It removes both the shared logical volume on which the filesystem resides and the associated stanza in the **/etc/filesystems** file.

To remove a shared filesystem:

1. On the source node, vary on the volume group, if needed, using the SMIT **varyonvg** fastpath.

   You can use the C-SPOC utility on a volume group that is varied off; however, you must specify the **-f** flag.
2. Enter the SMIT fastpath `cl_admin`

3. In SMIT, select **HACMP Logical Volume Management** > **Shared Filesystems > Remove a Shared Filesystem** and press Enter.
4. Press the F4 key to obtain a picklist of existing filesystems from which you may select one. Set the **Remove Mount Point** option to **yes** if you want to remove the mount point in the same operation. When you finish entering data, press Enter.

   The C-SPOC utility removes the filesystem on the source (local) node. The filesystem is *not* removed on remote nodes until the volume group on which the filesystem is defined is activated.
5. To check the status of the C-SPOC command execution on both nodes, view the C-SPOC log file in **/tmp/cspoc.log**.

Maintaining Physical Volumes

Administrative tasks that involve shared physical volumes include:

- Adding a Disk Definition to Cluster Nodes Using C-SPOC
- Removing a Disk Definition on Cluster Nodes Using C-SPOC
- Using SMIT to Replace a Cluster Disk
- Managing Data Path Devices with C-SPOC.

Adding a Disk Definition to Cluster Nodes Using C-SPOC

The nodes must be configured as part of the cluster.

To add a raw disk on selected cluster nodes:

1. Enter the fastpath `smitty cl_admin`
2. In SMIT, select **HACMP Physical Volume Management** > **Add a Disk to the Cluster** and press Enter. SMIT displays a list of nodes in the cluster and prompts you to select the nodes where the disk definition should be added.
3. Select one or more nodes where you want to have the new disk configured. The system generates a list of available disk types based on those disk types known to the first node in the list.
4. Select the type of disk you want to add to the cluster.

   The set of panels that follow depend on the disk type selected. Possible disk types include:

   - SCSI (various types listed)
   - SSA (available types listed)

Adding a SCSI Disk

If you select a SCSI disk type to define, SMIT displays a list of parent adapter/node name pairs. You are prompted to select one parent adapter per node.

To add a SCSI disk:

1. Select a parent adapter for each node connected to the disk and accept the selection.

SMIT displays the AIX 5L SCSI **Add a Disk** panel with all entries filled in except **Connection Address**:

Node Name(s) to which Disk is Attached	name of node(s) filled in.
Device Type	disk
Disk type	1000mb
Disk interface	scsi
Description	1.0GB SCSI Disk Drive
Node-Parent adapter Pairs	nodea:scsi0,nodeb:scsi1,...etc.
CONNECTION address	Use F4 to display the list; select the address.
ASSIGN physical volume identifier	**yes** is the default.

2. Select the connection address and press Enter.

   C-SPOC executes the necessary commands to define the disk on all selected nodes.

Adding an SSA Disk

If you select an SSA disk to define, SMIT displays the AIX 5L **Add an SSA Logical Disk** panel.

To add an SSA disk:

1. Select the connection address and other attributes as needed and press Enter:

Node Name(s) to which Disk is Attached	name of node(s) filled in.
Device Type	disk
Disk type	hdisk
Disk interface	ssa
Description	SSA Logical Disk Drive
Parent	ssar
CONNECTION address	Press F4 for a list; add the address.
Location Label	(*Optional*) Press F1 for information.
ASSIGN physical volume identifier	**yes** is the default.
RESERVE disk on open	**yes** is the default.

Queue depth	(*Optional*) Press F3 for information on ranges allowed. Press F1 for help.
Maximum coalesce	(*Optional*) Press F3 for information on ranges allowed. Press F1 for help.

C-SPOC runs the necessary commands to define the disk on all selected nodes.

Removing a Disk Definition on Cluster Nodes Using C-SPOC

Before removing a disk from the cluster using C-SPOC, check that the disk to be removed is *not* currently part of an existing volume group. If it is, use the C-SPOC **cl_reducevg** command to remove a physical volume from a volume group.

To remove a configured disk on all selected nodes in the cluster:

1. Enter the fastpath `smitty cl_admin`
2. In SMIT, select **HACMP Physical Volume Management** > **Remove a Disk from the Cluster** and press Enter.

   SMIT displays a list of nodes in the cluster that currently have the disk configured and prompts you to select the nodes where the disk should be removed.
3. Select the one or more node names from which you want to remove the disk configuration. (You may have removed the cable from some of the nodes in the cluster and only want the disk configuration removed from those nodes.)

   SMIT displays the AIX 5L **Remove a Disk** panel with the selected disks displayed.
4. For the entry **Keep the disk definition in database** select **yes t**o keep the definition in the database; select **no** to delete the disk from the database. Press Enter.

   C-SPOC sends the **rmdev** command to all nodes listed to remove the selected disk.

Using SMIT to Replace a Cluster Disk

The SMIT interface simplifies the process of replacing a failed disk. You can use this process with SCSI or SSA disks. Like the manual procedure for replacing a failed disk, SMIT uses C-SPOC commands.

Note: If you have VPATH devices configured, the procedure for replacing a cluster disk using C-SPOC requires additional steps. For instructions, see Replacing A Cluster Disk with a VPATH Device.

Prerequisites

Before you replace a disk, ensure that:

- You have root user privilege to perform the disk replacement.
- The volume group is mirrored.
- You have a replacement disk with an assigned PVID configured on all nodes in the resource group to which the volume group belongs. If you do *not* have the PVID assigned, run **chdev** on all nodes in the resource group.
- To add a new disk, remove the old disk and put the new one in its place.

Follow the steps in the section Adding a SCSI Disk or Adding an SSA Disk as appropriate. Ensure that the **ASSIGN physical volume identifier** field is set to **yes**.

Replacing a Cluster Disk

To replace a disk in the cluster:

1. Locate the failed disk. Make note of the PVID volume group.
2. Enter `smitty cl_admin`
3. In SMIT, select **HACMP Physical Volume Management** > **Cluster Disk Replacement** and press Enter.

   SMIT displays a list of disks that are members of volume groups contained in cluster resource groups. There must be two or more disks in the volume group where the failed disk is located. The list includes the volume group, the hdisk, the disk PVID, and the reference cluster node. (This node is usually the cluster node that has the volume group varied on.)

   Note: This process requires the volume group to be mirrored and the new disk that is available for replacement to have a PVID assigned to in on all nodes in the cluster. Use the **chdev** command to assign a PVID to the disk.

4. Select the disk for disk replacement (**source disk**) and press Enter.

   SMIT displays a list of those available shared disk candidates that have a PVID assigned to them, to use for replacement. (Only a disk that is of the same capacity or larger than the failed disk is suitable to replace the failed disk.)

5. Select the replacement disk (**destination disk**) and press Enter.

   SMIT displays your selections from the two previous panels.

6. Press Enter to continue or Cancel to terminate the disk replacement process.

   SMIT warns you that continuing will delete any information you may have stored on the destination disk.

7. Press Enter to continue or Cancel to terminate.

   SMIT displays a command status panel, and informs you of the **replacepv** recovery directory.

   If disk configuration fails and you wish to proceed with disk replacement, you must manually configure the destination disk. If you terminate the procedure at this point, be aware that the destination disk may be configured on more than one node in the cluster.

   The **replacepv** utility updates the volume group in use in the disk replacement process (on the reference node only).

   Note: SMIT displays the name of the recovery directory to use should **replacepv** fail. Make note of this information, as it is required in the recovery process.

   Configuration of the destination disk on all nodes in the resource group takes place.

8. If a node in the resource group fails to import the updated volume group, you must do this manually.

C-SPOC will *not* remove the failed disk information from the cluster nodes, hdisk, and pdisk. You must do this manually.

For information on recovering from a failed disk replacement, see the Cluster Disk Replacement Process Fails section in Chapter 3: Investigating System Components and Solving Common Problems in the *Troubleshooting Guide*.

Managing Data Path Devices with C-SPOC

All VPATH disk operations currently supported on AIX 5L are now supported by C-SPOC. You can define and configure VPATH devices, add paths, configure defined VPATHs, and remove VPATH devices. You can also display VPATH device and adapter configuration and status.

You must have SDD 1.3.1.3 or greater installed.

This section describes the following tasks:

- Displaying Data Path Device Configuration
- Displaying Data Path Device Status
- Displaying Data Path Device Adapter Status
- Defining and Configuring all Data Path Devices
- Adding Paths to Available Data Path Devices
- Configuring a Defined Data Path Device
- Removing a Data Path Device
- Converting ESS hdisk Device Volume Group to an SDD VPATH Device Volume Group
- Converting SDD VPATH Device Volume Group to an ESS hdisk Device Volume Group
- Replacing A Cluster Disk with a VPATH Device.

Displaying Data Path Device Configuration

To display data path device configuration:

1. Enter the fastpath `smit cl_admin`
2. In SMIT, select **HACMP Physical Volume Manager** > **Cluster Data Path Device Management > Display Data Path Device Configuration** and press Enter.

   SMIT displays the node picklist.
3. Select a node and press Enter.

   SMIT displays the configuration as shown in the following example for node `herbert`:

```
PVID: 000240bfd57e0746
herbert: vpath9 (Avail pv shvg1) 10612027 = hdisk59 (Avail ) hdisk65 (Avail )
PVID: 000240ffd5691fba
herbert: vpath12 (Avail ) 10C12027 = hdisk62 (Avail pv ) hdisk68 (Avail pv )
PVID: 000240ffd5693251
herbert: vpath14 (Avail pv ) 10E12027 = hdisk64 (Avail ) hdisk70 (Avail )
PVID: 000240ffd56957ce
herbert: vpath11 (Avail ) 10812027 = hdisk67 (Avail pv ) hdisk71 (Avail pv )
PVID: 0002413fef72a8f0
herbert: vpath13 (Avail pv ) 10D12027 = hdisk63 (Avail ) hdisk69 (Avail )
PVID: 0002413fef73d477
herbert: vpath10 (Avail pv ) 10712027 = hdisk60 (Avail ) hdisk66 (Avail )
```

Displaying Data Path Device Status

To display data path device status:

1. Enter the fastpath `smit cl_admin`
2. In SMIT, select **HACMP Physical Volume Manager** > **Cluster Data Path Device Management > Display Data Path Device Status** and press Enter.

   SMIT displays the node picklist.
3. Select a node and press Enter.
4. SMIT displays the status as shown in the following example for node `herbert`:

```
[TOP]
herbert: Total Devices : 6

PVID 000240bfd57e0746

herbert:
DEV#: 0 DEVICE NAME: vpath9 TYPE: 2105F20 SERIAL: 10612027
POLICY: Optimized
================================================================
Path# Adapter/Hard Disk State Mode Select Errors
0 fscsi1/hdisk59 OPEN NORMAL 1696 0
1 fscsi0/hdisk65 OPEN NORMAL 1677 0

PVID 000240ffd5691fba
[MORE...57]
```

Displaying Data Path Device Adapter Status

To display data path device adapter status:

1. Enter the fastpath `smit cl_admin`
2. In SMIT, select **HACMP Physical Volume Manager** > **Cluster Data Path Device Management > Display Data Path Device Adapter Status** and press Enter.

   SMIT displays the node picklist.
3. Select a node and press Enter.
4. SMIT displays the status as shown in the following example for node herbert:

```
herbert:

Active Adapters :2

Adpt#     Adapter Name   State     Mode     Select     Errors  Paths  Active
    0           fscsi1  NORMAL   ACTIVE       2204          0      6       1
    1           fscsi0  NORMAL   ACTIVE       2213          0      6       1
```

Defining and Configuring all Data Path Devices

To define and configure data path devices:

1. Enter the fastpath `smit cl_admin`
2. In SMIT, select **HACMP Physical Volume Manager** > **Cluster Data Path Device Management > Define and Configure all Data Path Devices** and press Enter.

   The command runs and the command status displays on the panel.

Adding Paths to Available Data Path Devices

To add paths to available data path devices:

1. Enter the fastpath `smit cl_admin`
2. In SMIT, select **HACMP Physical Volume Manager** > **Cluster Data Path Device Management > Add Paths to Available Data Path Devices** and press Enter.

   SMIT displays the list of node names.
3. Select one or more nodes, and press Enter.

   The command runs and the command status is displayed on the panel.

Configuring a Defined Data Path Device

To configure a defined data path device:

1. Enter the fastpath `smit cl_admin`
2. In SMIT, select **HACMP Physical Volume Manager** > **Cluster Data Path Device Management > Configure a Defined Data Path Device** and press Enter.

   SMIT displays the list of node names.
3. Select one or more nodes, and press Enter.

   SMIT displays the list of defined VPATHs by PVID.
4. Select a PVID and press Enter.

   The command runs and the command status is displayed on the panel.

Removing a Data Path Device

To remove a data path device:

1. Enter the fastpath `smit cl_admin`
2. In SMIT, select **HACMP Physical Volume Manager** > **Cluster Data Path Device Management > Remove a Data Path Device** and press Enter.

   SMIT displays the list of node names.
3. Select a node and press Enter.
4. Keep the definition in Data Base Selector.

   SMIT displays the list of devices.
5. Select one or more devices and press Enter.

   The command runs and the command status is displayed on the panel.

Converting ESS hdisk Device Volume Group to an SDD VPATH Device Volume Group

To convert ESS hdisk volume groups to SDD VPATH device volume groups:

1. Enter the fastpath `smit cl_admin`
2. In SMIT, select **HACMP Physical Volume Manager** > **Cluster Data Path Device Management > Convert ESS hdisk Device Volume Group to an SDD VPATH Device Volume Group** and press Enter.

SMIT displays the picklist of ESS hdisk volume groups.

3. Select the ESS hdisk volume group to convert and press Enter.

   SMIT displays the current resource group and volume group names.

4. Press Enter.

   The command runs and the command status is displayed on the panel.

Converting SDD VPATH Device Volume Group to an ESS hdisk Device Volume Group

To convert SDD VPATH device volume groups to ESS hdisk device volume groups:

1. Enter the fastpath `smit cl_admin`
2. In SMIT, select **HACMP Physical Volume Manager** > **Cluster Data Path Device Management > Convert SDD VPATH Device Volume Group to an ESS hdisk Device Volume Group** and press Enter.

   SMIT displays the picklist of SDD VPATH volume groups.

3. Select the one to convert and press Enter.

   SMIT displays the current resource group and volume group names.

4. Press Enter.
5. The command runs and the command status is displayed on the panel.

Replacing A Cluster Disk with a VPATH Device

If you need to replace a cluster disk that has a VPATH device configured, before you use C-SPOC, move the PVID of the VPATH devices to the corresponding hdisks. This is done by converting the volume group from VPATH devices to hdisks. After converting, use the C-SPOC procedure to replace a disk.

Note: The C-SPOC disk replacement utility does *not* recognize VPATH devices. If you do *not* convert the volume group from VPATH to hdisk, then during the C-SPOC disk replacement procedure, HACMP returns a “no free disks” message, although unused VPATH devices are available for replacement.

To replace a cluster disk that has a VPATH device configured:

1. Convert the volume group from VPATHs to hdisks. For instructions, see Converting SDD VPATH Device Volume Group to an ESS hdisk Device Volume Group.
2. Use the C-SPOC procedure to replace a cluster disk. For instructions, see Using SMIT to Replace a Cluster Disk.
3. Convert the volume group back to VPATH devices. For instructions, see Converting ESS hdisk Device Volume Group to an SDD VPATH Device Volume Group.

Configuring Cross-Site LVM Mirroring

This section describes the following tasks:

- Prerequisites

- Steps to Configure Cross-Site LVM Mirroring
- Showing and Changing Cross-Site LVM Mirroring Definition
- Removing a Disk from a Cross-Site LVM Mirroring Site Definition
- Troubleshooting Cross-Site LVM Mirroring.

Prerequisites

Before configuring cross-site LVM mirroring:

- Configure the sites and resource groups and run the HACMP cluster discovery process.
- Ensure that both sites have copies of the logical volumes and that **forced varyon** for a volume group is set to **Yes** if a resource group contains a volume group.

Steps to Configure Cross-Site LVM Mirroring

To configure cross-site LVM mirroring (cluster services are *not* running):

1. Enter the fastpath `smit cl_xslvmm`
2. Select **Add Disk/Site Definition for Cross-Site LVM Mirroring**.
3. Press F4 to see the list. Select a site from the list.
4. Select the physical volumes you want to assign to the selected site. Only disks connected to at least one node at two sites are displayed, as in the following example:

```
10002411f95594596     hdisk1 nodeA
20002411f95594595     hdisk2 nodeA
30002411f95594597     hdisk3 nodeA
40002411f95594595     hdisk4 nodeA
```

5. Repeat the steps to assign disks to the second site.
6. When you finish configuring the disks for cross-site LVM mirroring, create the shared volume groups that have the disks you included in the cross-site LVM mirror configuration.

   These volume groups should have **Enable Cross-Site LVM Mirroring** set to **True**. This setting causes verification to check the remote mirror configuration of the shared volume groups. See Creating a Shared Volume Group with C-SPOC for the procedure (or, Creating a Concurrent Volume Group on Cluster Nodes Using C-SPOC in Chapter 12: Managing Shared LVM Components in a Concurrent Access Environment if you are creating a concurrent volume group.
7. Add the volume groups to the appropriate resource groups. See Chapter 3: Configuring an HACMP Cluster (Standard).
8. Synchronize the cluster.
9. Start cluster services.
10. Add logical volumes to the volume groups. See Adding a Logical Volume to a Cluster Using C-SPOC or Adding a Concurrent Logical Volume to a Cluster in Chapter 12: Managing Shared LVM Components in a Concurrent Access Environment depending on the type of configuration (non-concurrent or concurrent).

Showing and Changing Cross-Site LVM Mirroring Definition

To display the current cross-site LVM mirroring definition:

1. Enter the fastpath `smit cl_xslvmm`
2. In SMIT, select **Change/Show Disk Site Definition for Cross-Site LVM Mirroring** and press Enter.
3. SMIT displays the current definition, as in the following example:

```
000000687fc7db9b      hdisk1  Node_Kiev_1             Site_1
 000000687fc7e1ce       hdisk2  Node_Kiev_1               Site_2
 00001281f551e35d       hdisk5  Node_Kiev_1               Site_2
 00001638ea883efc       hdisk3  Node_Kiev_1               Site_2
 000031994e4bbe3c       hdisk6  Node_Kiev_1               Site_1
 00004414d7723f69       hdisk7  Node_Kiev_1               Site_2
```

4. (*optional)* Change the current site assignments for disks.
5. Synchronize the cluster if you made any change.

Removing a Disk from a Cross-Site LVM Mirroring Site Definition

To remove a disk from the current cross-site LVM mirroring mapping:

1. Enter the fastpath `smit cl_xslvmm`
2. In SMIT, select **Remove a Disk from Site Definition for Cross-Site LVM Mirroring** and press Enter.
3. Press F4 for the list, then select the disks to remove from the list.

```
10002411f95594596     hdisk1 nodeA
20002411f95594595     hdisk2 nodeA
30002411f95594597     hdisk3 nodeA
40002411f95594595     hdisk4 nodeA
```

4. Verify the selections when SMIT asks if you are sure.
5. Synchronize the cluster.

Troubleshooting Cross-Site LVM Mirroring

If resynchronization fails, examine the AIX 5L **errorlog** for possible causes. You can also set up an AIX 5L error notification method.

Manual intervention is required in case of disk or disk subsystem failure:

- **Non-concurrent resource groups**. When the disks become available, the logical volume mirrors must be synchronized using the **Synchronize Shared LVM Mirrors** SMIT menu. Do this whether quorum was lost or *not*.
- **Concurrent resource groups**.
  - When the disks become available, the concurrent resource group must be brought ONLINE from the ERROR state if the quorum was lost during the disk failure. The mirrors will be synchronized automatically during the **rg_move** operation.
  - When the disks become available, the logical volume mirrors must be synchronized using the **Synchronize Concurrent LVM Mirrors** SMIT menu if the quorum was *not* lost, since in this case the resource group will *not* move to ERROR state.

To manually resynchronize the logical volume mirrors, use the appropriate SMIT menu: **Synchronize Concurrent LVM Mirrors** or **Synchronize Shared LVM Mirrors**.

Note: When you disable the cross-site mirror characteristic for a volume group, this turns off the verification check for the cross-site mirror configuration. It is *not* necessary to resynchronize after making this change.

Chapter 12: Managing Shared LVM Components in a Concurrent Access Environment

This chapter explains how to maintain shared LVM components in a concurrent access environment using the C-SPOC utility. This chapter includes specific procedures for managing volume groups, filesystems, logical volumes, physical volumes, and data path devices (VPATH disks).

When you add a concurrent capable volume group to a resource group, you can select the option to import the volume group onto all the destination nodes in the participating resource group. In addition, using SMIT, you can collect information about all volume groups available either on a local node or on all nodes in the cluster defined to the resource group, and later automatically import the appropriate volume groups to the destination nodes, if needed.

While the maintenance tasks for shared LVM components in concurrent access environments are similar to those of non-concurrent access environments, described in Chapter 11: Managing Shared LVM Components, there are some special considerations in a concurrent access environment.

You can use AIX 5L commands to do these procedures; however the C-SPOC utility is designed specifically to facilitate HACMP system management.

The main topics in this chapter include:

- Overview
- Understanding Concurrent Access and HACMP Scripts
- Maintaining Concurrent Access Volume Groups
- Maintaining Concurrent Volume Groups with C-SPOC
- Maintaining Concurrent Logical Volumes.

Overview

In a *concurrent access* environment, all cluster nodes can have simultaneous (concurrent) access to a volume group that resides on shared external disks.

Keep the following points in mind:

- When concurrent volume groups are created in AIX 5L, they are created as enhanced concurrent mode volume groups by default.

  In AIX 5L 5.2 or greater, you cannot create new SSA concurrent mode volume groups. If you have an SSA concurrent (mode=16) volume group that was created on AIX 5L 5.1, you can use it on AIX 5L 5.2. You can also convert these volume groups to enhanced concurrent mode.

If you are running AIX 5L 5.3, you *must* convert all volume groups to enhanced concurrent mode.

- If one node in a concurrent resource group runs a 64-bit kernel, enhanced concurrent mode must be used for that volume group.
- SSA concurrent mode is *not* supported on 64-bit kernels.
- SSA disks with the 32-bit kernel can use SSA concurrent mode.
- The C-SPOC utility does *not* work with RAID concurrent volume groups. You need to convert them to enhanced concurrent mode (otherwise, AIX 5L sees them an non-concurrent).

Note: You should convert both SSA and RAID concurrent volume groups to enhanced concurrent mode because AIX 5L enhanced concurrent mode provides improved functionality. For information about how to convert these types of volume groups to enhanced concurrent mode, see the section Converting Volume Groups to Enhanced Concurrent Mode.

For more information on enhanced concurrent mode, see Chapter 5 in the *Planning Guide*. Also, see Chapter 11: Managing Shared LVM Components, in this guide, the for information on Understanding Active and Passive Varyon in Enhanced Concurrent Mode and Enabling Fast Disk Takeover.

You can define concurrent access volume groups and logical volumes on all the disk devices supported by the HACMP software.

Note: You cannot define filesystems on a concurrent access volume group unless it is an enhanced concurrent mode volume group used as a serial resource.

The chapter first describes how the HACMP scripts handle concurrent access LVM components and then describes how to perform specific administrative tasks for LVM components in a concurrent access environment.

The final section of the chapter describes the **Physical Volume Management** HACMP SMIT panels, including how to manage VPATH disks.

Most maintenance tasks can be performed using the HACMP C-SPOC utility.

Understanding Concurrent Access and HACMP Scripts

You should seldom, if ever, need to intervene in a concurrent access cluster. In a concurrent access environment, as in a non-concurrent environment, the HACMP event scripts control the actions taken by a node and coordinate the interactions between the nodes. However, as a system administrator, you should monitor the status of the concurrent access volume groups when HACMP events occur.

When intervening in a cluster, you must understand how nodes in a concurrent access environment control their interaction with shared LVM components. For example, the HACMP **node_up_local** script may fail before varying on a volume group in concurrent mode. After

fixing whatever problem caused the script to fail, you may need to manually vary on the volume group in concurrent access mode. The following sections describe the processing performed by these scripts.

Nodes Join the Cluster

A node joining a cluster calls the **node_up_local** script, which calls the **cl_mode3** script to activate the concurrent capable volume group in concurrent access mode. If resource groups are processed in parallel, **process_resources** calls **cl_mode3.**

The **cl_mode3** script calls the **varyonvg** command with the **-c** flag. For more information about this command and its flags, see the section Activating a Volume Group in Concurrent Access Mode. If the concurrent capable volume group is defined on a RAID disk array device, the scripts use the **convaryonvg** command to vary on the concurrent volume groups in concurrent mode.

Nodes Leave the Cluster

Nodes leaving the cluster do *not* affect the concurrent access environment. They simply vary off the volume groups normally. The remaining nodes take no action to change the concurrent mode of the shared volume groups.

When a node has cluster services stopped with resource groups brought offline, it executes the **node_down_local** script, which calls the **cl_deactivate_vgs** script. This script uses the **varyoffvg** command to vary off the concurrent volume groups.

Maintaining Concurrent Access Volume Groups

The LVM enables you to create concurrent access volume groups that can be varied on in either concurrent access mode or non-concurrent access mode. Maintaining these volume groups may require you to perform any of the following tasks:

- Activating a Volume Group in Concurrent Access Mode
- Determining the Access Mode of a Volume Group
- Restarting the Concurrent Access Daemon (clvmd).
- Verifying a Concurrent Volume Group.

The following sections describe how to perform these tasks.

Tasks you can perform with C-SPOC are described in a section at the end of the chapter.

Activating a Volume Group in Concurrent Access Mode

As a system administrator, you may, at times, need to bring a resource group online. After correcting the failure, take the following steps to bring the resource group online:

1. Enter `smitty cl_admin`
2. In SMIT, select **HACMP Resource Group and Application Management > Bring a Resource Group Online**
3. Select the resource group to bring online and press Enter.

Activating Concurrent Access Volume Groups

To activate a volume group in concurrent access mode. Use the following procedure:

1. Enter `smit varyonvg`

   The **Activate a Volume Group** SMIT panel appears; it has an additional field in a concurrent access environment.

2. Enter the field values as follows:.

VOLUME GROUP name	Specify name of volume group.
RESYNCHRONIZE stale physical partitions?	Set this field to **no**.
Activate volume group in SYSTEM MANAGEMENT mode?	Accept the default (**no**).
FORCE activation of the Volume Group?	Accept the default (**no**).
Varyon VG in concurrent mode?	Set to **yes**.

3. Press Enter. The system prompts you to confirm. Press Enter again.

Determining the Access Mode of a Volume Group

To determine whether a volume group is a concurrent capable volume group and to determine its current mode, use the **lsvg** command specifying the name of the volume group as an argument. The **lsvg** command displays information about the volume group, as in the following example:

```
# lsvg vg1

VOLUME GROUP:   vg1                     VG IDENTIFIER:
00020adf00004c00000000f329382713
VG STATE:       active                  PP SIZE:        16 megabyte(s)
VG PERMISSION:  passive-only            TOTAL PPs:      542 (8672
megabytes)
MAX LVs:        256                     FREE PPs:       521 (8336
megabytes)
LVs:            3                       USED PPs:       21 (336
megabytes)
OPEN LVs:       0                       QUORUM:         2
TOTAL PVs:      1                       VG DESCRIPTORS: 2
STALE PVs:      0                       STALE PPs:      0
ACTIVE PVs:     1                       AUTO ON:        no
Concurrent:     Enhanced-Capable        Auto-Concurrent: Disabled
VG Mode:        Concurrent
Node ID:        2                       Active Nodes:   1 4
MAX PPs per PV: 1016                    MAX PVs:        32
LTG size:       128 kilobyte(s)         AUTO SYNC:      no
HOT SPARE:      no                      BB POLICY:      relocatable
```

To determine whether the volume group is concurrent capable, check the value of the **Concurrent** field. The volume group in the example was created as an Enhanced-capable volume group, as indicated by the value of this field. If this volume group was *not* a concurrent capable volume group, the value of this field would be *Non-Capable*.

To determine whether the volume group is activated in concurrent access mode, check the value of the **VG Mode** field. In the example, the volume group is activated in concurrent access mode. If this volume group had *not* been varied on in concurrent access mode, the value of this field would be *Non-Concurrent*.

The **Auto-Concurrent** field indicates whether the volume group should be varied on in concurrent access mode when the volume group is started automatically at system reboot. The value of this field is determined by the value of the **-x** option to the **mkvg** command when the volume group was created. In an HACMP environment, this option should always be disabled; HACMP scripts control when the volume should be varied on.

Restarting the Concurrent Access Daemon (clvmd)

As a system administrator, you may, at times, need to restart the concurrent access daemon used by SSA concurrent mode. The **clvmd** daemon normally gets started by the **varyonvg** command when you vary on a volume group in concurrent access mode by specifying the **-c** flag. You can restart the **clvmd** by re-executing the **varyonvg -c** command on the already varied-on concurrent access volume group. You cannot vary on an already varied-on volume group in a different mode, however.

You can also restart the **clvmd** daemon using the following SRC command:

```
startsrc -s clvmd
```

Enhanced concurrent mode uses the **gsclvmd** daemon, which starts when HACMP services are started. Beginning with HACMP 5.3, the Cluster Manager daemon explicitly controls the startup and shutdown of dependent software, such as RSCT and **gsclvmd.**

Note: In HACMP 5.3 and up, the Cluster Manager starts out of **inittab**.

Verifying a Concurrent Volume Group

On all nodes participating in a resource group that have the volume group defined, a volume group consistency check that HACMP runs during the verification process ensures the following:

- The **concurrent** attribute setting for the volume group is consistent across all related cluster nodes
- The list of PVIDs for this volume group is identical on all related cluster nodes
- An automatic corrective action of the cluster verification utility updates volume group definitions on all related cluster nodes
- Any problems detected are reported as errors.

Note: The SSA node number check is *not* performed for enhanced concurrent volume group that sit on the SSA hdisks. Disks that make up enhanced concurrent volume groups do *not* have any SSA-specific numbers assigned to them.

Maintaining Concurrent Volume Groups with C-SPOC

C-SPOC uses the AIX 5L CLVM capabilities that allow changes to concurrent LVM components without stopping and restarting the cluster.

See Chapter 11: Managing Shared LVM Components, for a general explanation of how C-SPOC works.

You can use the C-SPOC utility to do the following concurrent volume groups tasks:

- Create a concurrent volume group on selected cluster nodes (using hdisks or data path devices)
- Convert SSA concurrent or RAID concurrent volume groups to enhanced concurrent mode
- List all concurrent volume groups in the cluster
- Import a concurrent volume group
- Extend a concurrent volume group
- Reduce a concurrent volume group
- Mirror a concurrent volume group
- Unmirror a concurrent volume group
- Synchronize concurrent LVM mirrors by volume group.

Note: The volume group must be varied on in concurrent mode in order to do these tasks.

WARNING: If you have specified a forced varyon attribute for volume groups in a resource group, all nodes of the cluster must be available before making LVM changes. This ensures that all nodes have the same view of the state of the volume group. For more information on when it is safe to perform a forced varyon operation, and on instructions how to specify it in SMIT, see Forcing a Varyon of Volume Groups in Chapter 5: Configuring HACMP Resource Groups (Extended).

To perform concurrent resource group maintenance tasks, use the **HACMP System Management (C-SPOC) > HACMP Resource Group and Application Management** menu in SMIT.

This utility allows you to take a concurrent resource group online or offline (along with its resources—IP addresses, applications, and disks)—without stopping cluster services. For more information on Resource Group Migration, see Requirements before Migrating a Resource Group in Chapter 15: Managing Resource Groups in a Cluster.

Note: If you run a **ps** command during a C-SPOC LVM operation to verify what processes are running, you will see output similar to the following:

```
ps -ef | grep vg     root 11952 13522   0 08:56:25      -  0:00
ksh  /usr/es/sbin/cluster/cspoc/cexec cllvmcmd -extendvg -f
gdgpgogdhfhchgghdb gigegjhdgldbda.
```

That is a C-SPOC encapsulation of arguments/parameters when data is sent off to remote nodes.

Creating a Concurrent Volume Group on Cluster Nodes Using C-SPOC

Using C-SPOC simplifies the procedure for creating a concurrent volume group on selected cluster nodes. By default, the concurrent volume group will be created in enhanced concurrent mode. If you change this default to **false**, the disks must support concurrent mode (currently this is only true for SSA disks).

Note: For creating a concurrent volume path on VPATH disks, see Managing Data Path Devices with C-SPOC in Chapter 11: Managing Shared LVM Components. If you add a VPATH disk to a volume group made up of hdisks, the volume group will be converted to VPATHs on all nodes.

Before creating a concurrent volume group for the cluster using C-SPOC, check that:

- All disk devices are properly attached to the cluster nodes.
- All disk devices are properly configured on all cluster nodes and listed as available on all nodes.
- The cluster concurrent logical volume manager is installed.
- All disks that will be part of the volume group are concurrent capable.
- SSA disk subsystem have unique non-zero node numbers.

To create a concurrent volume group for a selected list of cluster nodes:

1. Enter the fastpath `smit cl_admin`
2. In SMIT, select **HACMP Concurrent Logical Volume Management** > **HACMP Concurrent Volume Groups** > **Create a Concurrent Volume Group** (or **Create a Concurrent Volume Group with Data Path Devices**) and press Enter.

   SMIT displays a list of cluster nodes.
3. Select one or more nodes from the list of cluster nodes and press Enter.

   The system correlates a list of all free concurrent-capable physical disks that are available to all nodes selected. (Free disks are those disks that currently are *not* part of a volume group and have PVIDs.) SMIT displays the list of free physical disks in a multi-picklist by PVIDs. If you are creating a volume group with data path devices, only disks capable of hosting them will be listed.
4. Select one or more PVIDs from the list and press Enter.

   SMIT displays the **cl_mkvg** panel with a major number inserted into the **Major Number** data field. The system determines this free major number; do *not* change it.
5. Enter field values as follows:

Node Names	Names of the selected nodes are displayed.
VOLUME GROUP name	Enter a name for this concurrent capable volume group.

Physical partition SIZE in megabytes	Accept the default.
Volume Group MAJOR NUMBER	The system displays the number C-SPOC has determined to be correct. **WARNING**: Changing the volume group major number may result in the command's inability to execute on a node that does *not* have that major number currently available. Check for a commonly available major number on all nodes before changing this setting.
Enhanced Concurrent Mode	The default is **true.** If you select **false**, the disks must support concurrent mode (SSA concurrent mode).
Enable/Disable a Concurrent Volume Group for Cross-Site LVM Mirroring	Set this field to **True** to enable data mirroring between sites. The default is **False**.

C-SPOC verifies communication paths and version compatibility and then runs the command on all the nodes you selected.

If cross-site LVM mirroring is enabled, that configuration is verified when you verify and synchronize. For more information on configuring cross-site LVM mirroring, see Configuring Cross-Site LVM Mirroring in Chapter 11: Managing Shared LVM Components.

Note: If the major number entered on the SMIT panel was *not* free at the time that the system attempted to make the volume group the command will display an error for the node that did *not* complete the execution and continue to the other nodes. At the completion of the command the volume group will *not* be active on any node in cluster.

Note: Verification issues an error if you run this command on some but *not all* nodes that own that resource group. All nodes in the same resource group must activate this resource group in enhanced concurrent mode.

Converting Volume Groups to Enhanced Concurrent Mode

It is highly recommended that you convert all concurrent volume groups, including RAID and SSA concurrent volume groups, to enhanced concurrent mode because:

- RAID concurrent volume groups must be converted to enhanced concurrent mode for C-SPOC to handle them.
- SSA concurrent volume groups can be varied on, but *not* created, in AIX 5L v.5.2 or greater.

Before converting a volume group to enhanced concurrent mode, ensure that the volume group:

- Is included in a resource group
- Is varied off everywhere in the cluster.

To convert a concurrent volume group to enhanced concurrent mode:

1. Use the **clRGmove** utility to take the resource group offline.
2. Enter the fastpath `smit cl_admin`
3. In SMIT, select **HACMP Concurrent Logical Volume Management** > **Concurrent Volume Groups** > **Set Characteristics of a Concurrent Volume Group > Convert a Concurrent Volume Group to Enhanced Concurrent Mode** and press Enter.

   SMIT displays the volume groups.
4. Select the volume group you want to convert.
5. Confirm the selected volume group is the one to convert.

   The volume group is converted to enhanced concurrent mode. If it does *not* meet the requirements for the conversion you will receive messages to that effect.
6. Use the **clRGmove** utility to bring the resource group back online.

Listing All Concurrent Volume Groups in the Cluster

To list all concurrent volume groups in the cluster:

1. On the source node, enter `smit cl_admin`
2. In SMIT, select **HACMP Concurrent Logical Volume Management > Concurrent Volume Groups > List all Concurrent Volume Groups** and press Enter.

   SMIT displays a message asking if you want to view only active concurrent volume groups.
3. Select **yes** to see a list of active concurrent volume groups only, or select **no** to see a list of all concurrent volume groups.

Importing a Concurrent Volume Group with C-SPOC

The physical volumes (**hdisks**) in the volume group must be installed, configured, and available.

To import a concurrent volume group using the C-SPOC utility:

1. On any cluster node that can own the concurrent volume group (is in the participating nodes list for the resource group), vary on the volume group in concurrent mode, using the SMIT **varyonvg** fastpath (if it is *not* varied on already).
2. On the source node, enter `smit cl_admin`
3. In SMIT, select **HACMP Concurrent Logical Volume Management > Concurrent Volume Groups > Import a Concurrent Volume Group** and press Enter.

   SMIT displays a list of volume groups.
4. Select a volume group and press Enter.

   SMIT displays a list of physical volumes.
5. Select a physical volume and press Enter.

SMIT displays the **Import a Concurrent Volume Group** panel. Values for fields you have selected are displayed.

6. Use the defaults or enter the appropriate values for your operation:

VOLUME GROUP name	The name of the volume group that you are importing.
PHYSICAL VOLUME name	The name of one of the physical volumes that resides in the volume group. This is the hdisk name on the reference node.
Reference node	The node from which the physical disk was retrieved.
Volume Group MAJOR NUMBER	If you are *not* using NFS, use the default (which is the next available number in the valid range). If you are using NFS, you must be sure to use the same major number on all nodes. Use the **lvlstmajor** command on each node to determine a free major number common to all nodes.
Make this VG concurrent capable?	The default is **no**. Change to **yes** for concurrent VGs.
Make default varyon of VG concurrent?	The default is **no**. Change to **yes** for concurrent VGs.

7. If this panel reflects the correct information, press Enter to import the concurrent volume group. All nodes in the cluster receive this updated information immediately.
8. If you did this task from a cluster node that does *not* need the concurrent volume group varied on, vary off the volume group on that node.

Extending a Concurrent Volume Group with C-SPOC

Note: If you add a VPATH disk to a volume group made up of hdisks, the volume group will be converted to VPATHs on all nodes.

The physical volumes (**hdisks**) being added to the volume group must be installed, configured, and available.

To add a physical volume to a concurrent volume group using C-SPOC:

1. On any cluster node that can own the volume group (is in the participating nodes list for the resource group), vary on the volume group, using the SMIT **varyonvg** fastpath (if it is *not* varied on already in concurrent mode).
2. On the source node, enter `smit cl_admin`
3. In SMIT, select **HACMP Concurrent Logical Volume Management > Concurrent Volume Groups > Set Characteristics of a Concurrent Volume Group > Add a Physical Volume to a Concurrent Volume Group** and press Enter.

   SMIT displays a list of volume groups.
4. Select a volume group and press Enter.

   SMIT displays a list of those physical volumes that have PVIDs assigned to them.

5. Select the PVIDs to add to the volume group and press Enter.
6. SMIT displays the **Add a Physical Volume to a Concurrent Volume Group** panel, with the following entries filled in:

Resource Group name	The cluster resource group to which this concurrent volume group belongs.
Volume Group name	The name of the volume group where hdisks are to be added.
Reference node	The name of the node where the hdisks are found.
Physical Volume names	The names of the hdisks to be added to the volume group.

7. If this panel reflects the correct information, press Enter to add the disks to the concurrent volume group. All nodes in the cluster receive this updated information.
8. If you did this task from a cluster node that does *not* need the concurrent volume group varied on, vary off the volume group on that node.

Enabling or Disabling Cross-Site LVM Mirroring

To enable or disable cross-site LVM mirroring of a concurrent volume group:

1. Enter the fastpath `smit cl_admin`
2. In SMIT, select **HACMP Concurrent Logical Volume Management > Concurrent Volume Groups > Set Characteristics of a Concurrent Volume Group > Enable/Disable a Concurrent Volume Group for Cross-site Mirroring** and press Enter.

   SMIT displays a list of volume groups.
3. Select a volume group from the list.
4. Enable or disable the field **Enable Cross-Site LVM Mirroring** (**True** or **False**) and press Enter. If you disable the cross-site mirror configuration, do this for each node that can own the volume group.
5. Synchronize the cluster if you are enabling cross-site LVM mirroring and cluster services are *not* running. (If sites are configured, you can only run verification while cluster services are *not* running.)

Note: For more information on configuring cross-site LVM Mirroring, see Configuring Cross-Site LVM Mirroring in Chapter 11: Managing Shared LVM Components.

Removing a Physical Volume from a Concurrent Volume Group with C-SPOC

The physical volumes (**hdisks**) in the volume group must be installed, configured, and available.

To remove a physical volume from a concurrent volume group using the C-SPOC utility:

1. On any cluster node that can own the concurrent volume group (is in the participating nodes list for the resource group), vary on the volume group in concurrent mode, using the SMIT **varyonvg** fastpath (if it is *not* varied on already).
2. On the source node, enter `smit cl_admin`
3. In SMIT, select **HACMP Concurrent Logical Volume Management > Concurrent Volume Groups > Set Characteristics of a Concurrent Volume Group > Remove a Physical Volume from a Concurrent Volume Group** and press Enter.

   SMIT displays a list of volume groups.
4. Select the desired volume group and press Enter.

   SMIT displays a list of physical volumes.
5. Pick a physical volume and press Enter.
6. SMIT displays the **Remove a Volume from a Concurrent Volume Group** panel, with the following entries filled in:

VOLUME GROUP name	The name of the volume group that you are reducing.
Reference node	The node from which the name of the physical disk was retrieved.
PHYSICAL VOLUME name	The name of the physical volume that you want to remove.This is the hdisk name on the reference node.

7. If this panel reflects the correct information, press Enter to reduce the concurrent volume group. All nodes in the cluster receive this updated information immediately (before “lazy update”).
8. If you did this task from a cluster node that does *not* need the concurrent volume group varied on, vary off the volume group on that node.

Mirroring a Concurrent Volume Group Using C-SPOC

The physical volumes (**hdisks**) in the volume group must be installed, configured, and available.

To mirror a concurrent volume group using the C-SPOC utility:

1. On any cluster node that can own the concurrent volume group (is in the participating nodes list for the resource group), vary on the volume group in concurrent mode, using the SMIT **varyonvg** fastpath (if it is *not* varied on already).
2. On the source node, enter `smit cl_admin`
3. In SMIT, select **HACMP Concurrent Logical Volume Management > Concurrent Volume Groups > Mirror a Concurrent Volume Group** and press Enter.

   SMIT displays a list of volume groups.
4. Select a volume group and press Enter.

   SMIT displays a list of physical volumes.
5. Pick a physical volume and press Enter.

SMIT displays the **Mirror a Concurrent Volume Group** panel. Values for fields you have selected are displayed.

6. Enter values in other fields as follows:

Resource Group Name	The name of the resource group to which this concurrent volume group belongs is displayed.
VOLUME GROUP name	The name of the volume group that you want to mirror is displayed.
Reference node	The node from which the name of the physical disk was retrieved is displayed.
PHYSICAL VOLUME names	The name of a physical volume on the volume group that you want to mirror.This is the hdisk name on the reference node.
Mirror sync mode	**Foreground** is the default. Other choices are **Background** and **No Sync.**
Number of COPIES of each logical partition	The default is **2**. You can also select **3**.
Keep Quorum Checking On?	The default is **no.** You can also select **yes.**
Create Exact LV Mapping?	The default is **no**.

7. If this panel reflects the correct information, press Enter to mirror the concurrent volume group. All nodes in the cluster receive this updated information.
8. If you did this task from a cluster node that does *not* need the concurrent volume group varied on, vary off the volume group on that node.

Unmirroring a Concurrent Volume Group Using C-SPOC

The physical volumes (**hdisks**) in the volume group must be installed, configured, and available.

To unmirror a concurrent volume group using the C-SPOC utility:

1. On any cluster node that can own the concurrent volume group (is in the participating nodes list for the resource group), vary on the volume group in concurrent mode, using the SMIT **varyonvg** fastpath (if it is *not* varied on already).
2. On the source node, enter `smit cl_admin`
3. In SMIT, select **HACMP Concurrent Logical Volume Management > Concurrent Volume Groups > Unmirror a Concurrent Volume Group** and press Enter.

   SMIT displays a list of volume groups.
4. Select a volume group and press Enter.

   SMIT displays a list of physical volumes.

5. Select a physical volume and press Enter.

   SMIT displays the **Unmirror a Concurrent Volume Group** panel, with the selected fields filled in.

6. For other fields, use the defaults or enter the appropriate values:

Resource Group Name	The name of the resource group to which this concurrent volume group belongs is displayed.
VOLUME GROUP name	The name of the volume group that you want to mirror is displayed.
Reference node	The node from which the name of the physical disk was retrieved is displayed.
PHYSICAL VOLUME names	The name of a physical volume on the volume group that you want to unmirror.This is the hdisk name on the reference node.
Number of COPIES of each logical partition	The default is **2**. You can also select **3**.

7. If this panel reflects the correct information, press Enter to unmirror the concurrent volume group. All nodes in the cluster receive this updated information.
8. If you did this task from a cluster node that does *not* need the concurrent volume group varied on, vary off the volume group on that node.

Synchronizing Concurrent Volume Group Mirrors

The physical volumes (**hdisks**) in the volume group must be installed, configured, and available.

To synchronize concurrent LVM Mirrors by volume group using the C-SPOC utility:

1. On any cluster node that can own the concurrent volume group (is in the participating nodes list for the resource group), vary on the volume group in concurrent mode, using the SMIT **varyonvg** fastpath (if it is *not* varied on already).
2. On the source node, enter `smit cl_admin`
3. In SMIT, select **HACMP Concurrent Logical Volume Management > Synchronize Concurrent LVM Mirrors > Synchronize By Volume Group** and press Enter.

   SMIT displays a list of volume groups.

4. Select a volume group and press Enter.

   SMIT displays a list of physical volumes.

5. Pick a physical volume and Press Enter.
6. SMIT displays the **Synchronize Concurrent LVM Mirrors by Volume Group** panel, with the chosen entries filled in.

7. For other fields, use the defaults or the appropriate values for your operation:

Resource Group Name	The name of the resource group to which this concurrent volume group belongs is displayed.
VOLUME GROUP name	The name of the volume group that you want to mirror is displayed.
Reference node	The node from which the name of the physical disk was retrieved is displayed.
Number of Partitions to Sync in Parallel	Set the range from 1 to 32.
Synchronize All Partitions	The default is **no.**
Delay Writes to VG from other cluster nodes during this Sync	The default is **no**.

8. If this panel reflects the correct information, press Enter to synchronize LVM mirrors by the concurrent volume group. All nodes in the cluster receive this updated information.
9. If you did this task from a cluster node that does *not* need the concurrent volume group varied on, vary off the volume group on that node.

Maintaining Concurrent Logical Volumes

You can use the C-SPOC utility to do many maintenance tasks on concurrent logical volumes. You can create a new logical volume or change the size of the logical volume. After you complete the procedure using SMIT, the other cluster nodes are updated with the new information.

Concurrent logical volumes tasks:

- List all concurrent logical volumes by volume group
- Add a concurrent logical volume to a volume group
- Remove a concurrent logical volume
- Make a copy of a concurrent logical volume
- Remove a copy of a concurrent logical volume
- Show the characteristics of a concurrent logical volume
- Synchronize concurrent LVM mirrors by logical volume
- Verify disk availability.

Listing All Concurrent Logical Volumes in the Cluster

To list all concurrent logical volumes in the cluster:

1. Enter `smit cl_admin`

2. In SMIT, select **HACMP Concurrent Logical Volume Management> Concurrent Logical Volumes > List all Concurrent Logical Volumes by Volume Groups** and press Enter.

   SMIT prompts you to confirm that you want to view only active concurrent volume groups.

3. Select **yes** to see a list of active concurrent volume groups only, or select **no** to see a list of all concurrent volume groups.

Adding a Concurrent Logical Volume to a Cluster

To add a concurrent logical volume to a cluster:

1. Enter the fastpath `smit cl_admin`
2. In SMIT, select **HACMP Concurrent Logical Volume Management** > **Concurrent Logical Volumes** > **Add a Concurrent Logical Volume** and press Enter.

   SMIT displays a list of concurrent volume groups.

3. Select a concurrent logical volume and press Enter.

   SMIT displays a list of physical volumes.

4. Select a physical volume and press Enter.

   The **Add a Concurrent Logical Volume** panel appears, with chosen fields filled in as shown in the sample below:

Resource Group name	ccur_rg
VOLUME GROUP name	concurrentvg1
Reference node	a1
*** Number of LOGICAL PARTITIONS**	[]
PHYSICAL VOLUME names	hdisk16
Logical volume NAME	If you are using this logical volume in a cross-site LVM mirroring configuration, name this logical volume appropriately, for example, *LV1site1*.
Logical volume TYPE	**jfs2** is recommended for cross-site LVM mirroring configurations.
POSITION on physical volume	middle
RANGE of physical volumes	minimum
MAXIMUM NUMBER of PHYSICAL VOLUMES to use for allocation	[]
Number of COPIES of each logical partition	Use at least two copies for cross-site LVM mirror configurations.

Mirror Write Consistency?	**no** (for concurrent environments)
Allocate each logical partition copy on a SEPARATE physical volume?	It is recommended to set this field to **super strict** if a forced varyon operation is specified for the volume groups. For a cross-site LVM mirror configurations, select either **superstrict** or **yes**.
RELOCATE the logical volume during reorganization	yes
Logical volume LABEL	[]
MAXIMUM NUMBER of LOGICAL PARTITIONS	[512]
Enable BAD BLOCK relocation?	no
SCHEDULING POLICY for writing logical partition copies	parallel
Enable WRITE VERIFY?	no
File containing ALLOCATION MAP	[]
Stripe Size?	[*Not* Striped]

The default logical volume characteristics are most common; however if you are using cross-site LVM mirroring, the characteristics recommended are noted above.

5. Make changes if necessary for your system and press Enter. Other cluster nodes are updated with this information.

Removing a Concurrent Logical Volume

To remove a concurrent logical volume on any node in a cluster:

1. Enter the fastpath `smit cl_admin`
2. In SMIT, select **HACMP Concurrent Logical Volume Management** > **Concurrent Logical Volumes** > **Remove a Concurrent Logical Volume** and press Enter.

   C-SPOC displays a list of concurrent logical volumes, organized by HACMP resource group.
3. Select the logical volume you want to remove and press Enter.
4. To check the status of the C-SPOC command execution on other cluster nodes, view the C-SPOC log file in **/tmp/cspoc.log**.

Setting Characteristics of a Concurrent Logical Volume

You can use C-SPOC to do the following tasks:

- Add copies to a concurrent logical volume

- Remove copies from a concurrent logical volume
- Show the current characteristics of a concurrent logical volume
- Synchronize the concurrent LVM mirrors by logical volume.

Adding a Copy to a Concurrent Logical Volume Using C-SPOC

To add a copy to a concurrent logical volume on all nodes in a cluster:

1. Enter the fastpath `smit cl_admin`
2. In SMIT, select **HACMP Concurrent Logical Volume Management** > **Concurrent Logical Volumes > Set Characteristics of A Concurrent Logical Volume > Add a Copy to a Concurrent Logical Volume** and press Enter.

   SMIT displays a list of logical volumes arranged by resource group.
3. Select a logical volume from the picklist and press Enter.

   SMIT displays a list of physical volumes.
4. Select a physical volume and press Enter.

   SMIT displays the **Add a Copy to a Concurrent Logical Volume** panel with the Resource Group, Logical Volume, Reference Node and default fields filled.
5. Enter the new number of mirrors in the **NEW TOTAL number of logical partitions** field and press Enter.

   The C-SPOC utility changes the number of copies of this logical volume on all cluster nodes.
6. To check the status of the C-SPOC command execution on all nodes, view the C-SPOC log file in **/tmp/cspoc.log**.

Removing a Copy from a Concurrent Logical Volume Using C-SPOC

To remove a copy of a concurrent logical volume on all nodes in a cluster:

1. Enter the fastpath `smit cl_admin`
2. In SMIT, select **HACMP Concurrent Logical Volume Management** >**Concurrent Logical Volumes >Set Characteristics of A Concurrent Logical Volume > Remove a Copy from a Concurrent Logical Volume** and press Enter.

   SMIT displays a list of logical volumes arranged by resource group.
3. Select a logical volume from the picklist and press Enter.

   SMIT displays a list of physical volumes.
4. Select the physical volumes from which you want to remove copies and press Enter.

   SMIT displays the **Remove a Copy from a Concurrent Logical Volume** panel with the Resource Group, Logical Volume name, Reference Node and Physical Volume names fields filled in.
5. Enter the new number of mirrors in the **NEW maximum number of logical partitions copies** field and check the **PHYSICAL VOLUME name(s) to remove copies from** field to make sure it is correct and press Enter.

   The C-SPOC utility changes the number of copies of this logical volume on all cluster nodes.

6. To check the status of the C-SPOC command execution on all nodes, view the C-SPOC log file in **/tmp/cspoc.log**.

Show Characteristics of a Concurrent Logical Volume Using C-SPOC

To show the current characteristics of a concurrent logical volume:

1. On the source node, vary on the volume group in concurrent mode, using the SMIT **varyonvg** fastpath.
2. Enter the fastpath `smit cl_admin`
3. In SMIT, select **HACMP Concurrent Logical Volume Management** >**Concurrent Logical Volumes > Show Characteristics of A Concurrent Logical Volume** and press Enter.

   SMIT displays a list of logical volumes arranged by resource group.
4. Select a logical volume from the picklist and press Enter.

   SMIT displays the **Show Characteristics of A Concurrent Logical Volume** panel with the **Resource Group**, and **Logical Volume name** fields filled in. The **Option list** field offers three choices for the display.
5. Select **Status**, **Physical Volume map**, or **Logical Partition map** in the **Option list** field and press Enter.

   SMIT displays the characteristics of a concurrent logical volume in the selected format.

Synchronizing LVM Mirrors by Logical Volume

The physical volumes (**hdisks**) in the volume group must be installed, configured, and available.

To synchronize concurrent LVM mirrors by logical volume:

1. On any cluster node that can own the concurrent volume group (is in the participating nodes list for the resource group), vary on the volume group, using the SMIT **varyonvg** fastpath (if it is *not* varied on already).
2. On the source node, enter `smit cl_admin`
3. In SMIT, select **HACMP Concurrent Logical Volume Management > Synchronize Concurrent LVM Mirrors > Synchronize By Logical Volume** and press Enter.

   SMIT displays a list of logical volumes.
4. Select a logical volume and press Enter.

   SMIT displays a list of physical volumes.
5. Select a physical volume and press Enter.

   SMIT displays the **Synchronize LVM Mirrors by Volume Group** panel, with the chosen entries filled in.
6. Enter values in other fields as needed for your operation:

Resource Group Name	The name of the resource group to which this logical volume belongs is displayed.

LOGICAL VOLUME name	The name of the logical volume that you want to synchronize is displayed.
Reference node	The node from which the name of the physical disk was retrieved is displayed.
Number of Partitions to Sync in Parallel	Set the range from 1 to 32.
Synchronize All Partitions	The default is **no.**
Delay Writes to VG from other cluster nodes during this Sync	The default is **no**.

7. If this panel reflects the correct information, press Enter to synchronize LVM mirrors by the concurrent logical volume.

   All nodes in the cluster receive this updated information.

8. If you did this task from a cluster node that does *not* need the concurrent volume group varied on, vary off the volume group on that node.

Chapter 13: Managing the Cluster Topology

This chapter describes how to reconfigure the cluster topology.

The main topics include:

- Reconfiguring a Cluster Dynamically
- Dynamic Cluster Topology Changes
- Viewing the Cluster Topology
- Configuring Communication Interfaces/Devices to the Operating System on a Node
- Swapping IP Addresses between Communication Interfaces Dynamically
- Replacing a PCI Hot-Pluggable Network Interface Card
- Changing a Cluster Name
- Changing the Configuration of Cluster Nodes
- Changing the Configuration of an HACMP Network
- Changing the Configuration of Communication Interfaces
- Managing Persistent Node IP Labels
- Changing the Configuration of a Global Network
- Changing the Configuration of a Network Module
- Changing the Configuration of a Site
- Synchronizing the Cluster Configuration.

Reconfiguring a Cluster Dynamically

When you configure an HACMP cluster, configuration data is stored in HACMP-specific object classes in the Configuration Database (ODM). The AIX 5L ODM object classes are stored in the default system configuration directory (DCD), **/etc/es/objrepos**.

You can make certain changes to both the cluster topology and to the cluster resources while the cluster is running. This is called a dynamic reconfiguration (DARE). You can make a combination of resource and topology changes via one dynamic reconfiguration operation.

All dynamic topology configuration changes allowed in an HACMP configuration are now supported in HACMP/XD configurations that include clusters with sites defined. This includes changes to XD-type (HAGEO or GLVM) networks, interfaces, sites, nodes, and NIM values. HACMP handles the resource groups that have primary and secondary instances (running on nodes at different sites) properly during these dynamic reconfiguration changes. See Chapter 14: Managing the Cluster Resources for information on supported dynamic changes to resources and resource groups.

If you have dependent resource groups in the cluster, see the section Reconfiguring Resources in Clusters with Dependent Resource Groups in Chapter 14: Managing the Cluster Resources for information on making dynamic reconfiguration changes to the cluster topology.

Note: No automatic corrective actions take place during a DARE.

At cluster startup, HACMP copies HACMP-specific ODM classes into a separate directory called the Active Configuration Directory (ACD). While a cluster is running, the HACMP daemons, scripts, and utilities reference the Configuration Database data stored in the active configuration directory (ACD) in the HACMP Configuration Database.

If you synchronize the cluster topology or cluster resources definition while the Cluster Manager is running on the local node, this action triggers a dynamic reconfiguration event. In a dynamic reconfiguration event, the HACMP Configuration Database data in the Default Configuration Directories (DCDs) on all cluster nodes is updated and the HACMP Configuration Database data in the ACD is overwritten with the new configuration data. The HACMP daemons are refreshed so that the new configuration becomes the currently active configuration.

The dynamic reconfiguration operation (that changes *both* resources and topology) progresses in the following order:

- Releases any resources affected by the reconfiguration
- Reconfigures the topology
- Acquires and reacquires any resources affected by the reconfiguration operation.

Requirements before Reconfiguring

Before making changes to a cluster definition, ensure that:

- HACMP is installed on all nodes.
- All nodes are up and running HACMP and able to communicate with each other: *no node may be in a forced down state.*
- The cluster is stable; no recent event errors or config_too_long messages exist.

Synchronizing Configuration Changes

When you change the topology or the resources of a cluster, you update the data stored in the HACMP Configuration Database in the DCD. For example, when you add an additional network interface to a cluster node, you must add the interface to the cluster definition so that the cluster nodes can recognize and use it.

When you change the cluster definition on one cluster node, you must also update the HACMP Configuration Databases on the other cluster nodes, a process called *synchronization*. Synchronization causes the information stored in the DCD on the local cluster node to be copied to the HACMP Configuration Database object classes in the DCD on the other cluster nodes.

When synchronizing the cluster triggers a dynamic reconfiguration event, HACMP verifies that both cluster topology and cluster resources are correctly configured, even though you may have only changed an element of one of these. Since a change in topology may invalidate the resource configuration, and vice versa, the software checks both.

Dynamic Cluster Topology Changes

DARE (Dynamic Reconfiguration) supports resource and topology changes done in one operation. For information on DARE operations in clusters with dependent resource groups, see Reconfiguring Resources in Clusters with Dependent Resource Groups.

You can make the following changes to the cluster topology in an active cluster, dynamically:

- Adding or removing nodes
- Adding or removing network interfaces
- Swapping a network interface card
- Changing network module tuning parameters
- Adding a new Heartbeating over Aliasing network
- Changing an active network to (but *not* from) Heartbeating over Aliasing
- Changing the address offset for a Heartbeating over Aliasing network
- Adding, changing or removing a network interface or a node configured in a Heartbeating over Aliasing network.

All topology configuration changes allowed in an HACMP configuration are now supported in HACMP/XD configurations. Supported changes include changes for XD-type networks, interfaces, sites, nodes, and NIM values. During the dynamic reconfiguration changes, HACMP properly handles resource groups that contain replicated resources (groups that have primary and secondary instances running on nodes at different sites).

To avoid unnecessary processing of resources, use **clRGmove** to move resource groups that will be affected by the change before you make the change. When dynamically reconfiguring a cluster, HACMP will release resource groups if this is found to be necessary, and they will be reacquired later. For example, HACMP will release and reacquire the resource group that is using the associated service IP address on a network interface affected by the change to topology.

Viewing the Cluster Topology

When you view the cluster topology, you are viewing the HACMP Configuration Database data stored in the DCD, *not* the data stored in the ACD.

Note: You can use WebSMIT to view graphical displays of sites, networks, nodes and resource group dependencies.

For more information on WebSMIT, see Chapter 2: Administering a Cluster Using WebSMIT.

Before making changes to a cluster topology, view the current configuration.

To view the cluster topology:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Topology Configuration > Show HACMP Topology** and press Enter.

SMIT displays the panel with the following options. Each option provides a different view of the cluster.

3. Select the appropriate option for the task at hand:

SMIT option	Description
Show Cluster Topology	Provides complete information about the cluster topology including the nodes in the cluster, their interfaces, and the networks that connect them.
Show Cluster Definitions	Lists the names of all clusters accessible from this node.
Show Topology Information by Node	Provides information about cluster nodes and their interfaces.
Show Topology Information by Networks	Provides information about the networks that connect the cluster nodes.
Show Topology Information by Communication Interface	Lists the network interfaces defined in the cluster.
Extended Configuration > Extended Topology Configuration > Configure HACMP Sites > Change/Show a Site	Provides information about the sites defined in the cluster.

Using the cltopinfo Command

You can also use the **/usr/es/sbin/cluster/utilities/cltopinfo** command to view the cluster topology configuration. See the man page or the description in Appendix C: HACMP for AIX Commands. The command shows all the topology information and you can choose to see it organized by node, network, or interface.

Managing Communication Interfaces in HACMP

This section describes the options under the **System Management (C-SPOC) > HACMP Communication Interface Management** SMIT menu:

- Configuring Communication Interfaces/Devices to the Operating System on a Node
- Updating HACMP Communication Interfaces/Devices with AIX 5L Settings
- Swapping IP Addresses between Communication Interfaces Dynamically
- Hot-Replacing a PCI Network Interface Card.

Configuring Communication Interfaces/Devices to the Operating System on a Node

You can configure communication interfaces/devices to AIX 5L without leaving HACMP SMIT, by using the **System Management (C-SPOC) > HACMP Communication Interface Management** SMIT path.

To configure communication interfaces/devices to the operating system on a node:

1. Enter the fastpath `smit hacmp`
2. In SMIT, select **System Management (C-SPOC) > HACMP Communication Interface Management > Configure Communication Interfaces/Devices to the Operating System on a Node** and press Enter.

   A picklist with node names appears.
3. Select a node on which to configure a network interface or device from the picklist.
4. Select a communication interface or a device type and press Enter:

Network Interfaces	This option leads to the AIX 5L configuration SMIT menus for a particular node. Each network interface must be defined to the operating system before it can be used by HACMP. It is equivalent to running **smitty mktcpip**.
RS232 Devices	This option leads to the AIX 5L configuration SMIT menus for a particular node. Each TTY device must be defined to the operating system before it can be used by HACMP. It is equivalent to running **smitty tty**.
Target-Mode SCSI Devices	This option leads to the AIX 5L configuration SMIT menus for a particular node. Each target-mode SCSI device must be defined to the operating system before it can be used by HACMP. It is equivalent to running **smitty scsia**.
Target-Mode SSA Devices	This option leads to the AIX 5L configuration SMIT menus for a particular node. Each target-mode SSA device must be defined to the operating system before it can be used by HACMP. It is equivalent to running **smitty ssaa**.
X.25 Communication Interfaces	This option leads to the AIX 5L configuration SMIT menus for a particular node. Each X.25 Communication Interface device must be defined to the operating system before it can be used by HACMP.
SNA Communication Links	This option leads to the AIX 5L configuration SMIT menus for a particular node. Each SNA Communication Link device must be defined to the operating system before it can be used by HACMP.

Physical Disk Devices	This option leads to the AIX 5L configuration SMIT menus for a particular node. Each physical disk device must be defined to the operating system before it can be used by HACMP.

5. To finish configuring the communication interface or device on a node, fill in the fields in the corresponding AIX 5L SMIT panel that will open. For instructions, see the *AIX 5L System Administration Guide*.

Updating HACMP Communication Interfaces/Devices with AIX 5L Settings

When you define communication interfaces/devices by entering or selecting an HACMP IP label or device, HACMP discovers the associated AIX 5L network interface name. HACMP expects this relationship to remain unchanged. If you change the IP Label/Address associated with the AIX 5L network interface after configuring and synchronizing the cluster, HACMP will *not* function correctly.

If this problem occurs, you can reset the network interface IP Label/Address with the AIX 5L settings using the SMIT HACMP **System Management (C-SPOC**) menu.

Use this SMIT selection to update HACMP after you make any changes to the underlying AIX 5L configuration of the mapping of a network interface to an IP Label/Address. For example, you should update HACMP after modifying the **nameserver** or **/etc/hosts**.

You must stop cluster services, make the change, and then restart cluster services to apply it to the active configuration. You cannot make these changes dynamically.

To update HACMP with new AIX 5L settings:

1. Stop cluster services on the node where you are running the update.
2. Enter `smit hacmp`
3. In SMIT, select **System Management (C-SPOC) > HACMP Communication Interface Management > Update HACMP Communication Interfaces Communication Interface with Operating System Settings** and press Enter.

   A picklist with node names appears.
4. Select a node on which to run the utility and press Enter.

   The update automatically calls commands to explicitly re-populate the HACMPadapter Configuration Database with the updated entries and then explicitly re-syncs the HACMPadapter class only.
5. Start cluster services.

Swapping IP Addresses between Communication Interfaces Dynamically

As a systems administrator, you may at some point experience a problem with a network interface card on one of the HACMP cluster nodes. If this occurs, you can use the dynamic communications interface swap feature to swap the IP address of an active service communication interface with the IP address of another active, available communication interface on the same node and network. Cluster services do *not* have to be stopped to perform the swap.

You can use this feature to move an IP address off of a NIC that is behaving erratically without shutting down the node. It can also be used if a hot pluggable communication device is being replaced on the node. Hot pluggable NICs can be physically removed and replaced without powering off the node.

This feature can also be used to move the persistent IP label to another network interface.

If hardware address swapping is enabled, the hardware address will be swapped along with the IP address.

Restrictions on IP Address Swapping

Note the following restrictions:

- The dynamic communications interface swap is *not* allowed for service IP labels that are configured on networks using IP aliases.
- The dynamic communications swap feature is *not* supported on the SP switch network.
- The dynamic IP address swap can only be performed within a single node. To move an IP address to another node, move its resource group using the **clRGmove** Resource Group Management utility. See Chapter 15: Managing Resource Groups in a Cluster.
- The dynamic IP address Swap IP function is *not* supported for IP addresses on XD_data networks (these networks are used in clusters with sites that run HACMP/XD for GLVM or HAGEO). If you perform this action on an XD_data network, the cluster may run an incorrect **network_down** event, which could generate an incorrect **rg_move** event. If there is no takeover node available for the resource group, then it may be moved into the OFFLINE state, rendering the resources unavailable.

Procedure for Swapping an IP Address Dynamically

Make sure that no other HACMP events are running before swapping a network interface.

To dynamically swap an IP address between communication interfaces:

1. Enter `smit hacmp`
2. In SMIT, select **System Management (C-SPOC) > HACMP Communication Interface Management > Swap IP Addresses Between Communication Interfaces** and press Enter.

   SMIT displays a list of available service interfaces. It also displays those interfaces that have persistent labels placed on them, but are *not* hosting service IP labels. This allows you to move the persistent label to another interface.
3. Select the service communication interface you want to remove from cluster use, and press Enter.

   SMIT displays a list of available non-service interfaces.
4. Select a non-service interface and press Enter.

   The **Swap IP Addresses Between Communication Interfaces** menu appears.
5. Verify the service IP label, and the non-service IP label you have chosen. If this is correct, press Enter.

   SMIT prompts you to confirm that you want to do this operation.
6. Press Enter *only* if you are sure you want to swap the communication interface.

After the swapping of IP addresses between communication interfaces, the service address becomes an available non-service interface. At this point, you can take action to repair the faulty network interface card. If you have a hot pluggable network interface card, you can replace it while the node and cluster services are up. Otherwise, you will have to stop cluster services and power down the node to replace it.

If you have a hot pluggable network interface card, HACMP makes the interface unavailable when you pull it from the node. When the new card is placed in the node, the network interface is incorporated into the cluster as an available non-service IP label again. You can then use the dynamic network interface swap feature again to swap the IP address back to the original network interface.

If you need to power down the node to replace the faulty network interface card, HACMP will configure the service and non-service addresses on their original communication interfaces when cluster services are restarted. You do *not* need to use the dynamic network interface swap feature again to swap the interfaces. HACMP does *not* record the swapped interface information in the AIX 5L Configuration Database (ODM). Therefore, the changes are *not* persistent across system reboots or cluster restarts.

Replacing a PCI Hot-Pluggable Network Interface Card

This section takes you through the process of replacing a PCI hot plug network interface card.

Special Considerations

Keep the following in mind before you replace a hot-pluggable PCI network interface card:

- Be aware of the following consideration: If a network interface you are hot-replacing is the only available keepalive path on the node where it resides, *you must shut down HACMP on this node in order to prevent a partitioned cluster* while the interface is being replaced.
- SMIT gives you the option of stopping cluster services on this node with resource groups brought offline. From this point, you can manually hot-replace the network interface card.
- Hot-replacement of Ethernet, Token-Ring, FDDI, and ATM network interface cards is supported. This process is *not* supported for non-IP communication devices.
- You should manually record the IP address settings of the network interface being replaced to prepare for unplanned failures.
- You should *not* attempt to change any configuration settings while the hot replacement is in progress.
- To avoid a network failure when using multiple dual-port Ethernet adapter cards on the same node for a particular network, you must configure the interfaces on different physical dual-port Ethernet adapter cards.

  Note: Hot-replacement of the dual-port Ethernet adapter used to configure two interfaces for one HACMP IP network is currently *not* supported

Hot-Replacing a PCI Network Interface Card

The SMIT interface simplifies the process of replacing a hot-pluggable PCI network interface card. HACMP supports only one PCI hot plug network interface card replacement via SMIT at one time per node.

Note: If the network interface was alive before the replacement process began, then between the initiation and completion of the hot-replacement, the interface being replaced is in a maintenance mode. During this time, network connectivity monitoring is suspended on the interface for the duration of the replacement process.

Scenario 1 (Live NICs Only)

Follow the procedure below when hot-replacing the following:

- A live PCI network service interface in a resource group and with an available non-service interface
- A live PCI network service interface *not* in a resource group and with an available non-service interface
- A live PCI network boot interface with an available non-service interface.

1. Go to the node on which you want to replace a hot-pluggable PCI network interface card.
2. Type `smit hacmp`
3. In SMIT, select **System Management (C-SPOC) > HACMP Communication Interface Management > PCI Hot Plug Replace a Network Interface Card** and press Enter.

   SMIT displays a list of available PCI network interfaces that are hot-pluggable.
4. Select the network interface you wish to hot-replace. Press Enter. The service address of the PCI interface is moved to the available non-service interface.
5. SMIT prompts you to physically replace the network interface card. After you have replaced the card, you are asked to confirm that replacement has occurred.

   If you select **yes**, the service address will be moved back to the network interface that has been hot-replaced. On aliased networks, the service address will *not* move back to the original network interface, but will remain as an alias on the same network interface. The hot-replacement is complete.

   If you select **no**, you must manually reconfigure the interface settings to their original values:

   a. Run the **drslot** command to take the PCI slot out of the removed state.

   b. Run **mkdev** on the physical interface.

   c. Use **ifconfig** manually as opposed to `smit chinet`, **cfgmgr**, or **mkdev** in order to avoid configuring duplicate IP addresses or an unwanted boot address.

Scenario 2 (Live NICs Only)

Follow the procedure below when hot-replacing a live PCI network service interface on a resource group but with no available non-service interface

1. Go to the node on which you want to replace a hot-pluggable PCI network interface card.
2. Enter `smit hacmp`
3. Select **System Management (C-SPOC) > HACMP Communication Interface Management > PCI Hot Plug Replace a Network Interface Card** and press Enter.

   SMIT displays a list of available PCI network interfaces that are hot-pluggable.

4. Select the network interface you wish to hot-replace and press Enter.

   SMIT prompts you to choose whether to move the resource group to another node during the replacement process in order to ensure its availability.

5. If you choose to do this, SMIT gives you the option of moving the resource group back to the node on which the hot-replacement took place after completing the replacement process.

   If you do *not* move the resource group to another node, it will be offline for the duration of the replacement process.

6. SMIT prompts you to physically replace the card. After you have replaced the network interface card, you are asked to confirm that replacement has occurred.

   If you select **Yes**, the hot-replacement is complete.

   If you select **No**, you must manually reconfigure the interface settings to their original values:

   a. Run the **drslot** command to take the PCI slot out of the removed state.

   b. Run **mkdev** on the physical interface.

   c. Use **ifconfig** manually as opposed to `smit chinet`, **cfgmgr**, or **mkdev** in order to avoid configuring duplicate IP addresses or an unwanted boot address.

   d. (*If applicable*) Move the resource group back to the node from which you moved it in step 5.

Scenario 3 (Non-alive NICs Only)

Follow the procedure below when hot-replacing the following:

- A non-alive PCI network service interface in a resource group and with an available non-service interface
- A non-alive PCI network service interface *not* in a resource group and with an available non-service interface
- A non-alive PCI network boot interface with an available non-service interface.

1. Go to the node on which you want to replace a hot-pluggable PCI network interface card.
2. Enter `smit hacmp`
3. Select **System Management (C-SPOC) > HACMP Communication Interface Management > PCI Hot Plug Replace a Network Interface Card** and press Enter.

   SMIT displays a list of available PCI network interfaces that are hot-pluggable.

4. Select the network interface you wish to hot-replace. Press Enter.

   SMIT prompts you to physically replace the network interface card.

5. After you have replaced it, SMIT prompts you to confirm that replacement has occurred.

   If you select **yes**, the hot-replacement is complete.

   If you select **no**, you must manually reconfigure the interface settings to their original values:

   a. Run the **drslot** command to take the PCI slot out of the removed state.

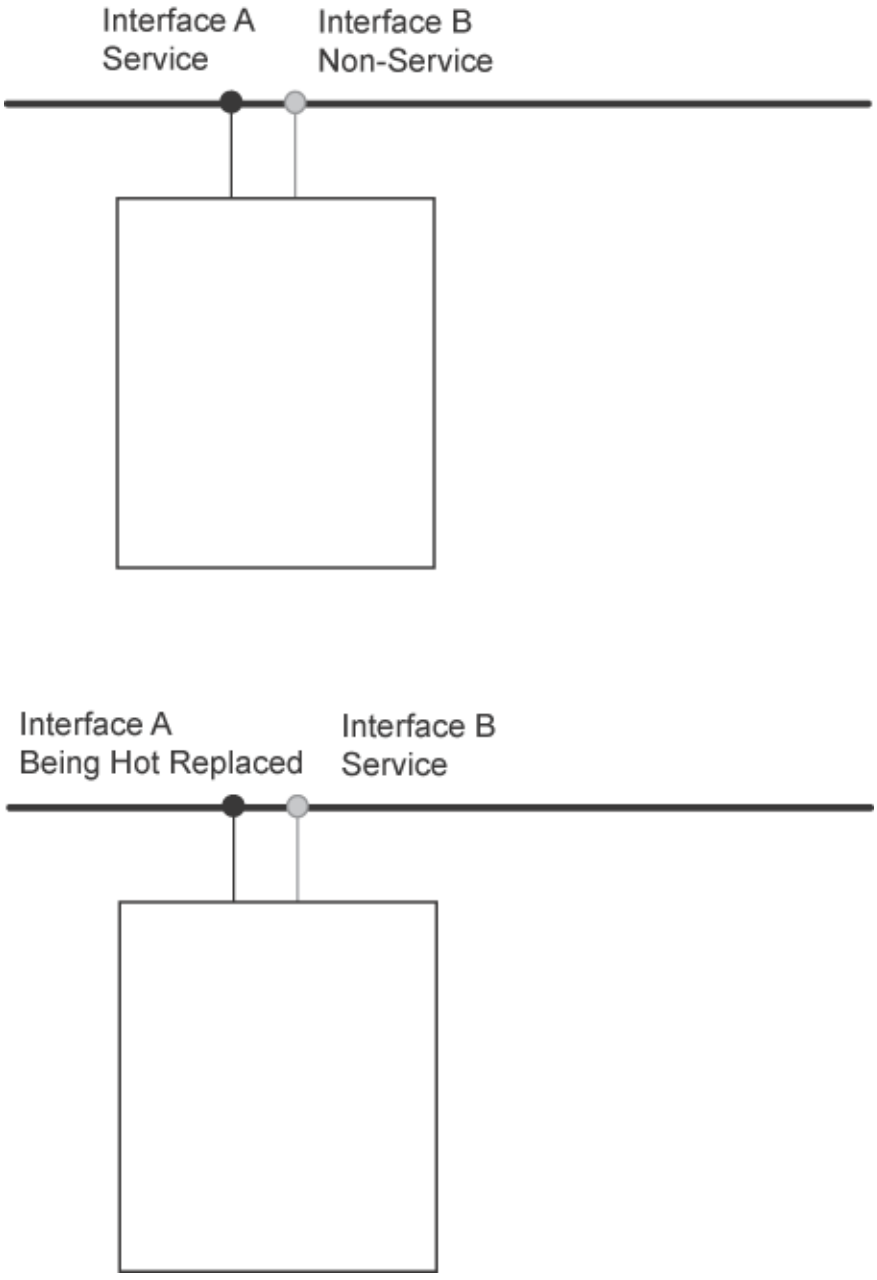
b. Run **mkdev** on the physical interface.

c. Use **ifconfig** manually as opposed to `smit chinet`, **cfgmgr**, or **mkdev** in order to avoid configuring duplicate IP addresses or an unwanted boot address.

Service Interface Failure During Hot-Replacement

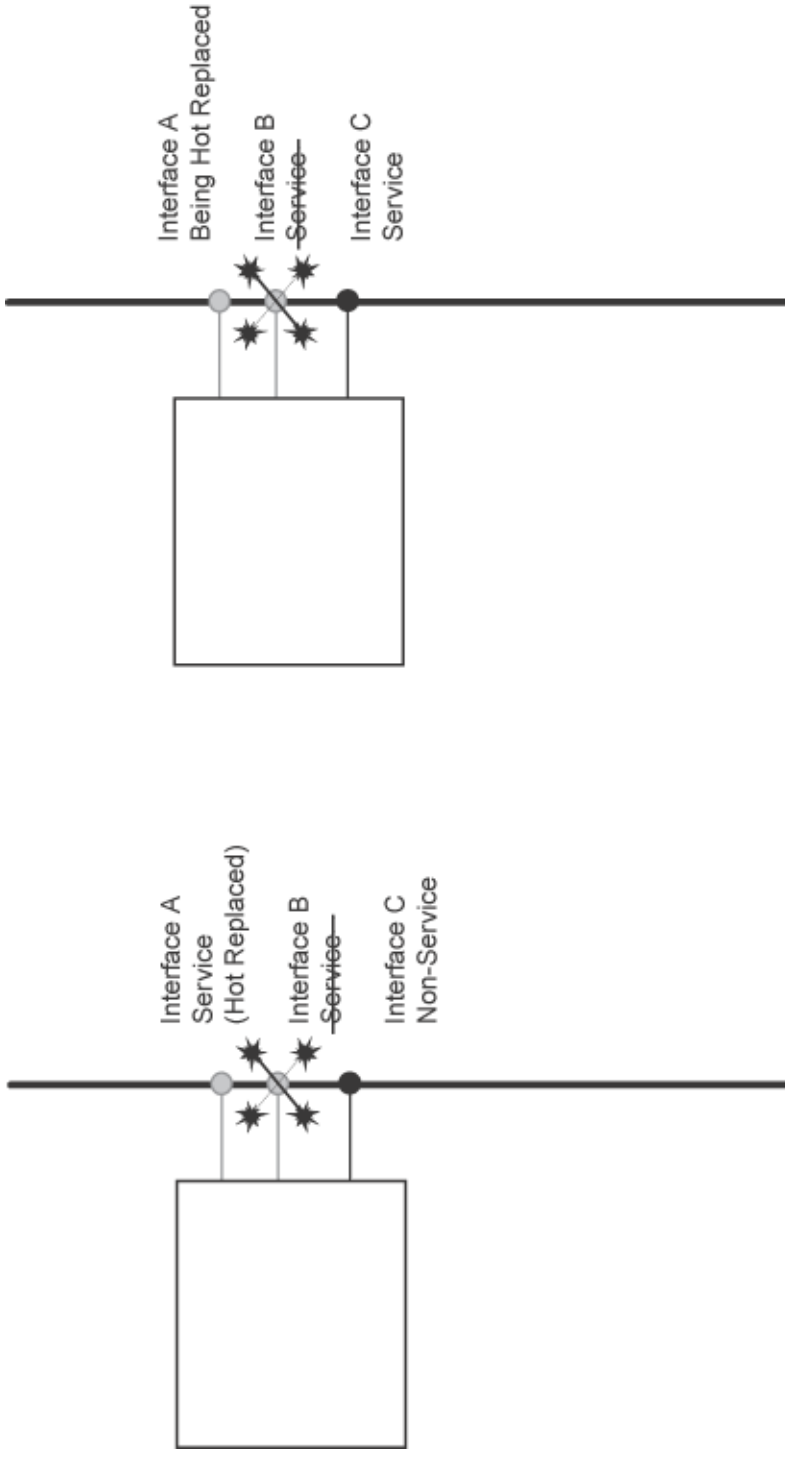
While an interface is unavailable during its replacement, HACMP continues processing events that occur during this time.

Consider, for example, where a node in a cluster has a service interface (Interface A) and an available non-service interface (Interface B) on the same network. If you want to hot-replace Interface A, the service network address will first be swapped to Interface B.



Behavior of Interface B while Interface A Being Hot-Replaced

Now consider that Interface B (now the service interface) fails while the hot-replace of Interface A is in progress. If there is another available non-service interface (C), HACMP does a swap of Interface B to Interface C. When the hot-replacement is finished, the service network settings are swapped from Interface C back to Interface A (the newly replaced interface), and Interface C is reconfigured to non-service settings.



Behavior of Interface C while Interface A Being Hot-Replaced and Interface B Fails

If there are no extra available non-service interfaces, then between the time Interface B (the service interface) fails and the replacement of Interface A is complete, the node has no network connectivity on that network. In this case, if there are no other sufficient network paths alive on

the node for keepalive traffic, a partitioned cluster results. If there are sufficient other network paths alive for keepalive traffic, then a local network failure event is generated for the network to which Interfaces A and B belong.

Any resource group dependent on a service interface in that same network moves to another node, thus the service address moves with the resource group. Following hot plug replacement, Interface A (the newly replaced interface) is reconfigured to a non-service address *not* currently used on that node and network.

Hot-Replacing an ATM Network Interface Card

ATM network interface cards support multiple logical interfaces on one network interface card. An ATM network interface hot- replacement is managed the same as other network interface cards, with the following exceptions:

- All logical interfaces configured on the card being replaced that are *not* configured for and managed by HACMP are lost during the replacement process. They will *not* be reconfigured on the newly replaced ATM entered interface card. All other logical interfaces configured for and managed by HACMP on the ATM network interface card being replaced are restored when the replacement is complete.
- Since it is possible to have more than one service interface configured on an ATM network interface card—thus multiple resource groups on one ATM network interface—when you hot-replace an ATM network interface card, SMIT leads you through the process of moving each resource group on the ATM interface, one at a time.

Recovering from PCI Hot Plug Network Interface Card Failure

If an unrecoverable error causes the hot-replacement process to fail, HACMP may be left in a state where your network interface is unconfigured and still in maintenance mode. To recover from this, manually fix the script, then run `smit clruncmd` to remove any maintenance modes that are still set. You can also use **ifconfig** to reconfigure the network settings of the interface.

Changing a Cluster Name

When changing the name of a cluster, you must stop cluster services, make the change, and then restart cluster services to apply it to the active configuration. You cannot make these changes dynamically.

To change a cluster's name:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Topology Configuration > Configure an HACMP Cluster > Add/Change/Show an HACMP Cluster**, and press Enter.

   SMIT displays the cluster definition with the current value for the cluster name filled in.
3. Enter the name change. A cluster name can include alphabetic and numeric characters and underscores; it cannot begin with a numeric. Use no more than 32 characters.

4. After the command completes, return to the HACMP SMIT menus to perform further topology reconfiguration or to synchronize the changes you made. To synchronize the cluster topology, return to the **Extended Configuration** panel and select the **Extended Verification and Synchronization** option.

Changing the Configuration of Cluster Nodes

As the system administrator of an HACMP cluster, you may need to perform any of the following tasks relating to cluster nodes:

- Adding one or more nodes to the cluster
- Removing a node from the cluster
- Changing the attributes of a cluster node.

Adding a Cluster Node to the HACMP Configuration

You can add a node to an active cluster dynamically. You do *not* need to stop and restart cluster services on the already-participating cluster nodes for the new node to become part of the cluster.

Take the following steps on any active cluster node (called the local node from here on), to add the new node to the cluster topology definition:

1. Enter `smit hacmp`.
2. In SMIT, select **Extended Configuration > Extended Topology Configuration > Configure HACMP Nodes > Add a Node to the HACMP Cluster** and press Enter.

   SMIT displays the **Add a Node to the HACMP Cluster** panel.
3. Enter the name of the node (or nodes) that you want to add to the cluster. A node name can include alphabetic and numeric characters and underscores, but cannot have a leading numeric. Use no more than 32 characters. Separate multiple names with spaces. If you specify a duplicate node name, the operation fails. Press Enter to add the node or nodes to the cluster definition.
4. Optionally, you can add a Communication Path. Press the F4 key to see the picklist that displays the contents of **/etc/hosts**. Enter one resolvable IP Label/Address (may be the hostname), IP address, or Fully Qualified Domain Name for the node. This path will be taken to initiate communication with the node. Examples are: "`NodeA`", "`10.11.12.13`", and "`NodeC.ibm.com`".
5. After the command completes, return to the HACMP SMIT menus to perform further topology reconfiguration or to synchronize the changes you made. To synchronize the cluster, return to the **Extended Configuration** panel and select the **Extended Verification and Synchronization** option (You can wait and synchronize after doing the resource configuration. if you prefer).
6. On the newly added node, start cluster services to integrate it into the cluster.

Adding Nodes to a Resource Group

Once you have added the new node to the cluster topology, you can continue by adding the new node (or nodes) to the list of participating nodes in a resource group.

In a non-concurrent resource group with the startup policy of either Online on Home Node Only or Online on First Available Node, if you give the new node the highest priority by specifying it first in the list of participating nodes, the newly added node will acquire control of the resource group when you start up cluster services on this node. This can be useful when you want the new node to take over a specific resource. For example, you may be adding a high-powered node to a cluster that runs a heavily used database application and you want this application to run on the newly added node.

WARNING: When adding a node to a cluster with a resource group that has *disk fencing enabled*, add the node to the concurrent resource group immediately. All nodes in a concurrent access cluster must participate in the concurrent access resource group. Include the new node in this resource group immediately to avoid the possibility of unrecoverable data loss.

When you are finished adding the node to a resource group:

1. Synchronize the cluster. Return to the **Extended Configuration** panel and select the **Extended Verification and Synchronization** option.
2. When you press Enter, the cluster resources are dynamically reconfigured.

Removing a Cluster Node from the HACMP Configuration

You can remove a node from an active cluster dynamically. However, before removing a node from the cluster, you must remove the node from any resource groups it participates in and synchronize resources.

To remove a cluster node:

1. Stop cluster services on the node to be removed (usually this is done by stopping cluster services with the **Move Resource Groups** option).
2. On another active node, enter `smit hacmp`
3. In SMIT, select **Extended Configuration > Extended Topology Configuration > Configure HACMP Nodes > Remove a Node in the HACMP Cluster**. SMIT displays a list of all cluster nodes.
4. Select the node you want to remove and press Enter. SMIT prompts you to confirm that you want to proceed. Press Enter again to remove the node from the cluster.

   Note: When you remove a node from the cluster topology, all communication path information associated with the node is also removed, its resources are released and reacquired, and the node is removed from the resource configuration.

5. On the local node, return to the **Extended Configuration** panel and select the **Extended Verification and Synchronization** option to synchronize the cluster. When the synchronization completes, the node is removed from the cluster definition.

Changing the Name of a Cluster Node

When changing the name of a cluster node, you must stop cluster services, make the change, and then restart cluster services to apply it to the active configuration.

To change the name of a cluster node:

1. Enter `smit hacmp`
2. Select the following options: **Extended Configuration > Extended Topology Configuration > Configure HACMP Nodes > Change/Show a Node in the HACMP Cluster** and press Enter.

   SMIT displays a picklist of cluster nodes.
3. Make your selection and press Enter.

   SMIT displays the current node name.
4. Enter the new name for the node in the **New Node Name** field. A node name can include alphabetic and numeric characters and underscores, but cannot have a leading numeric. Use no more than 32 characters. When you finish entering data, press Enter. SMIT makes the changes you specified.
5. After the command completes, return to the HACMP SMIT menus to perform further topology reconfiguration or to synchronize the changes you made. To synchronize the cluster topology, return to the **Extended Configuration** panel and select the **Extended Verification and Synchronization** option.

   The change is propagated through both the cluster topology and resource configuration.

Changing the Configuration of an HACMP Network

As the system administrator of an HACMP cluster, you may need to perform any of the following tasks relating to cluster networks:

- Adding a Network
- Changing Network Attributes
- Removing an HACMP Network
- Converting an HACMP Network to use IP Aliasing
- Establishing Default and Static Routes on Aliased Networks
- Converting an SP Switch Network to an Aliased Network
- Disabling IPAT via IP Aliases
- Controlling Distribution Preferences for Service IP Label Aliases.

Adding a Network

See Configuring HACMP Networks and Heartbeat Paths in Chapter 4: Configuring HACMP Cluster Topology and Resources (Extended).

Changing Network Attributes

You can change attributes of both IP and non-IP networks. You cannot change network attributes dynamically.

Changing the Network Attribute to Private for Oracle Inter-Node Communication

ORACLE uses the **private** network attribute setting to select networks for Oracle inter-node communications. This attribute is *not* used by HACMP and will *not* affect HACMP in any way.The default attribute is **public**.

Changing the network attribute to **private** makes the network Oracle-compatible by changing all interfaces to service (as well as changing the attribute in HACMPnetwork ODM).

To configure private networks for use by Oracle:

1. Configure the network and add all interfaces. You cannot change the attribute if the network has no interfaces.
2. Change the network attribute to **private**. See Steps for Changing an IP-Based Network below.
3. Private networks must have *either* all boot *or* all service interfaces. If the network has all boot interfaces (the default when using discovery) HACMP converts these interfaces to service. (Oracle only looks at service interfaces.)
4. Synchronize the cluster after changing the attribute.

Note: Once you define the network attribute as **private** you cannot change it back to **public**. You have to delete the network and then redefine it to HACMP. (It defaults to **public**.)

Steps for Changing an IP-Based Network

To change the name or an attribute of an HACMP network:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Topology Configuration > Configure HACMP Networks > Change/Show a Network in the HACMP Cluster** and press Enter.
3. Select the IP-based network to change. (The following panels depend on the type of network you selected to change.)
4. SMIT displays the **Change/Show a Network in the HACMP Cluster** panel. You can change the name and the mechanism for configuring the IP address on the network interface assigned to this network in the corresponding fields:

Network Name	The current name of the network is displayed.
New Network Name	The new name for this network. Use no more than 32 alphanumeric characters and underscores; do *not* use a leading numeric.
Network Type	Listed according to the chosen network.

Netmask	The netmask of the selected network is displayed, for example, 255.255.255.0.
Enable IP Address Takeover via IP Aliases	The value in this field determines the mechanism by which an IP Address will be configured onto a network interface. By default, if the network and selected configuration supports adding an IP Alias to a network interface, it is set to **Yes**. Otherwise, it is **No**. If you explicitly want to use IPAT via IP Replacement, set this field to **No**. IP Replacement is the mechanism by which one IP address is first removed from, and then another IP address is added to, a single network interface. Note that this field is set to **No** by default after a migration of an SP Switch network that was configured to use IPAT via IP Replacement in HACMP.
IP Address Offset for Heartbeating over IP Aliases	The base address of a private address range for heartbeat addresses, for example 10.10.10.1. HACMP will use this address to automatically generate IP addresses for heartbeating for each boot interface in the configuration. Refer to the *Planning Guide* and your planning worksheet for more information on selecting a base address for use by Heartbeating over IP Aliases. Clear this entry to use the default heartbeating method.
Network Attribute	**public** is the default. Use **private** for Oracle.

5. Press Enter to change the definition of the network.
6. On the same node, synchronize the cluster configuration.

Steps for Changing Serial Devices

To change an attribute of a serial device:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Topology Configuration > Configure HACMP Networks > Change/Show a Network in the HACMP Cluster** and press Enter.

   SMIT displays a list of serial devices.
3. Select the serial device to change. (The following panels depend on the type of network you selected to change.)
4. Make the changes in the fields on the **Change/Show a Serial Device in the HACMP Cluster** panel as follows:

Network Name	The current name of the network is displayed.
New Network Name	The new name for the network.

Network Type	Valid types are **RS232**, **tmssa**, **tmscsi**, **diskhb**.

5. Press Enter to change the definition of the network.
6. On the same node, synchronize the cluster configuration.

Removing an HACMP Network

Note: Deleting all network interfaces associated with a network deletes the network definition from HACMP.

To remove a network from the HACMP cluster definition:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Topology Configuration > Configure HACMP Networks > Remove a Network from the HACMP Cluster** and press Enter.

   SMIT displays the **Select a Network to Remove** panel.
3. Select the network to remove.

   SMIT displays `Are you sure?`
4. Press Enter to remove the network. All of this network's subnets and their interfaces are removed from the HACMP configuration.
5. On the same node, synchronize the cluster configuration.

   If the Cluster Manager is running on the local node, the synchronization triggers a dynamic reconfiguration event. See Synchronizing the Cluster Configuration for more information.

Converting an HACMP Network to use IP Aliasing

If you want to change the cluster network configuration to use IPAT via Aliases instead of the previous IPAT via IP Replacement scheme for a specific network in the cluster, you should stop the cluster services on all nodes to make the change. This change is *not* allowed during a dynamic reconfiguration (DARE) of cluster resources.

Note: If you have an SP Switch network that has been configured in your cluster in HACMP prior to version 5.1, and want to convert the SP Switch to use the IP aliasing in HACMP, see the section Converting an SP Switch Network to an Aliased Network.

To convert an HACMP network to use IP Aliasing:

1. Stop the cluster services on all cluster nodes.
2. Verify that no HACMP interfaces are defined with HWAT on that network.
3. Verify that the network is configured to support gratuitous ARP in HACMP, by checking the **Extended Configuration > Extended Topology Configuration > Configure an HACMP Network Module** > **Show a Network Module** SMIT panel for the **Gratuitous ARP** setting for that network type.

4. To change the cluster network to use IPAT via IP Aliases instead of IPAT via IP Replacement, see the steps in this chapter in the section Changing Network Attributes.
5. Verify and synchronize the cluster.
6. Restart the cluster services.

For more information on IPAT via IP Aliases see the relevant chapters in the *Concepts and Facilities Guide* and in the *Planning Guide*.

Establishing Default and Static Routes on Aliased Networks

If you are setting up or converting to an IP aliased network and require establishing the default route, and possibly, other static routes that have to be established on the IP aliased service subnet, these routes will fail to be established automatically when the **rc.net** file runs at boot time. This is because there is no address on that subnet in the Configuration Database.

To ensure that these routes are established at boot time, we recommend that you also configure a persistent address on that subnet. After configuring the persistent address, HACMP configures the routes.

If you do *not* configure persistent addresses, then you should use your own scripts that will configure routes on aliased service subnets. For more information on the **rc.net** file see Chapter 1: Administering an HACMP Cluster and Chapter 9: Starting and Stopping Cluster Services.

Converting an SP Switch Network to an Aliased Network

When you migrate your cluster to HACMP 5.4 from a version prior to HACMP 5.1, your previously configured SP Switch network configuration remains valid. However, after migration, HACMP by default treats your network as non-aliased, although in reality it functions as an aliased network. Therefore, you may consider reconfiguring the existing SP Switch network configuration.

To change the cluster configuration to use IPAT via IP Aliases instead of the standard IPAT via IP Replacement scheme for the SP Switch network, stop the cluster services on all nodes and make the following changes to the communication interface definitions. Such changes are *not* allowed during a dynamic reconfiguration (DARE) of cluster resources.

To convert the SP Switch network to an aliased network, perform the following steps on all cluster nodes:

1. Stop the cluster services.
2. Enter `smit hacmp`
3. In SMIT, select **Extended Configuration > Extended Topology Configuration > Configure HACMP Communication Interfaces/Devices** > **Remove Communication Interfaces/Devices** and press Enter.
4. Remove the non-service interfaces on the SP Switch network.
5. Remove the boot label that you had previously configured within HACMP for your SP switch network.

6. Go back to the **Extended HACMP Verification and Synchronization** option and synchronize the cluster. If the cluster network meets all the requirements of an aliased network, the following message appears: "`Setting attribute for network <name> to use IP Aliasing for IP address takeover`".
7. In SMIT, select the **Extended Configuration > Extended Topology Configuration > Configure HACMP Communication Interfaces/Devices** and configure the base interface address on the SP Switch network as a boot IP label/address in HACMP.
8. Put the service IP-based communication interface on a *different* subnet than the boot interface to avoid errors during the verification process. If you have multiple service addresses they should all be on a different subnet than the boot interface.
9. Verify that HWAT is disabled for all communication interfaces on this network.
10. Verify that the network is configured to support gratuitous ARP in HACMP, by checking the **Gratuitous ARP** setting for that network type. See instructions in the section Changing the Tuning Parameters to Custom Values.
11. In the **Change/Show a Network in the HACMP Cluster** SMIT panel, set the **Enable IP Address Takeover via IP Aliases field** to **Yes** for this network.
12. Synchronize the cluster. HACMP verifies the configuration.
13. Restart the cluster services.

For more information on the SP Switch considerations, see Chapter 3: Planning Cluster Network Connectivity in the *Planning Guide*.

For more information on IPAT via IP Aliases see *Concepts and Facilities* and Chapter 3 in the *Planning Guide*.

Disabling IPAT via IP Aliases

If the network supports gratuitous ARP, you can configure the network in the HACMP cluster to use IPAT via IP Aliases during fallover.

There are subtle differences between the operation of a network using IP aliasing and one that does *not*. If you need to troubleshoot problems with external network equipment, clients, or applications, you may want to disable IPAT via IP Aliases on the network and use IPAT via IP Replacement instead.

To disable IPAT via IP Aliases facility for the entire network type:

1. In the SMIT **Change/Show a Network in the HACMP Cluster** panel, set the **Enable IP Address Takeover via IP Aliases field** to **No** for this network.
2. Change the service IP labels to be on different subnets.
3. Press Enter to accept the changes.
4. Synchronize the cluster configuration.

Controlling Distribution Preferences for Service IP Label Aliases

To control the placement of the service IP label aliases on the cluster node physical network interface cards, you can configure a distribution preference for the aliases of the service IP labels that are placed under HACMP control.

See the section Distribution Preference for Service IP Label Aliases: Overview in Chapter 4: Configuring HACMP Cluster Topology and Resources (Extended).

Changing the Configuration of Communication Interfaces

As a system administrator, you may need to perform any of the following tasks relating to cluster network interfaces:

- Configuring Multiple Logical Interfaces on the Same ATM NIC
- Adding HACMP Communication Interfaces/Devices
- Removing a Communications Interface from a Cluster Node.

Configuring Multiple Logical Interfaces on the Same ATM NIC

You can configure multiple logical network interfaces as HACMP communication interfaces, where all logical interfaces belong to the same physical ATM NIC, and each is defined as Classic IP or LANE. The cluster behaves as if it were configured on the same set of logical interfaces and each interface type is defined on a separate ATM NIC.

For more information on this functionality, refer to Chapter 3 in the *Planning Guide*.

Adding HACMP Communication Interfaces/Devices

You can add a network communication interface to an active cluster dynamically. You do *not* need to stop and restart cluster services for the network communication interface to become part of the cluster.

1. On the node getting the new network interface card, complete the prerequisite tasks:
   - Install the new network interface card.
   - Configure the new logical network interface to AIX 5L.
2. On all cluster nodes, update the **/etc/hosts** file to include the IP address of the new network interface.
3. On any cluster node, add the HACMP communication interface to the cluster topology definition.
4. Synchronize the cluster.

Adding a Communication Interface to an IP-Based Network

See Configuring Communication Interfaces/Devices to HACMP in Chapter 4: Configuring HACMP Cluster Topology and Resources (Extended).

Adding a Communication Device to a Non IP-Based Network

See Configuring Communication Interfaces/Devices to HACMP in Chapter 4: Configuring HACMP Cluster Topology and Resources (Extended).

Changing Communication Interface/Device Attributes

You cannot change the attributes of a communication interface or device dynamically. You must stop and restart cluster services to make the changed configuration the active configuration.

To change a communication interface or serial device for the cluster:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Topology Configuration > Configure HACMP Communication Interfaces/Devices** > **Change/Show HACMP Communication Interfaces/Devices** and press Enter.
3. Select the IP communication interface or the serial device from the picklist.

   Attributes for a communication interface include:

Node Name	The name of the node on which this network interface physically exists.
Network Interface	The network interface associated with the communication interface.
IP Label/Address	The IP label/address associated with this communication interface that will be configured on the network interface when the node boots. The picklist filters out IP labels/addresses already configured to HACMP.
Network Type	The type of network media/protocol (Ethernet, Token Ring, fddi, ATM) Select the type from the predefined list of network types.
Network Name	A unique name for this logical network.

   Attributes for a serial device are as follows:

Node Name	Define a node name for all serial service devices.
Device Name	Enter a device file name. • RS232 serial devices must have the device file name **/dev/tty*n***. • Target mode SCSI serial devices must have the device file name **/dev/tmscsi*n***. • Target mode SSA devices must have the device file name **/dev/tmssa*n***. For disk heartbeating, any disk device in an enhanced concurrent volume group is supported. It could be an **hdisk** or **vpath**, for example **/dev/hdisk*n***. *n* = the number of the device.
Device Path	For example, **/dev/tty0**

Network Type	This field is automatically filled in (RS232, tmssa, tmscsi, or diskhb) depending on the device name.
Network Name	This field is automatically filled in.

4. Press Enter after making the change (such as a new network name). HACMP now checks the validity of the configuration. You may receive warnings if a node cannot be reached.
5. Return to the **Extended Configuration** menu and select the **Extended Verification and Synchronization** option. If the configuration is verified and synchronized, proceed to the next step.
6. Restart cluster services.

   The change is propagated through the cluster. Cluster resources are modified, as specified.

Removing a Communications Interface from a Cluster Node

You can remove an HACMP communications interface from an active cluster dynamically; you do *not* need to stop and restart cluster services.

Note: Deleting all communications interfaces associated with a network deletes the network from HACMP.

To remove a communications interface from a cluster node:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Topology Configuration > Configure HACMP Communication Interfaces/Devices** > **Remove HACMP Communication Interfaces/Devices** and press Enter.
3. Select the IP communication interface(s) or the serial device(s) from the picklist and press Enter.

   When you remove a communications interface/device, all information associated with the interface/device is removed from the Configuration Database. SMIT prompts you to confirm that you want to do this operation. Press Enter again *only* if you are sure you want to remove the interface/device and its associated information.
4. On the same node, synchronize the cluster. If the Cluster Manager is running on the local node, the synchronization triggers a dynamic reconfiguration event. See Synchronizing the Cluster Configuration for more information.

   When the synchronization completes, the selected communications interfaces/devices are removed from the cluster topology definition.

Managing Persistent Node IP Labels

This section describes the following tasks:

- Configuring Persistent Node IP Labels/Addresses
- Changing Persistent Node IP Labels
- Deleting Persistent Node IP Labels.

Configuring Persistent Node IP Labels/Addresses

To configure persistent node IP labels on a specified node:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Topology Configuration > Configure HACMP Persistent Node IP Labels/Addresses** > **Add a Persistent Node IP Label/Address** and press Enter.
3. Select a cluster node.
4. Enter the field values as follows:

Node Name	The name of the node on which the IP Label/Address will be bound.
Network Name	The name of the network on which the IP Label/Address will be bound.
Node IP Label/Address	The IP Label/Address to keep bound to the specified node.

5. Press Enter. The resulting SMIT panel displays the current node name and persistent node IP labels defined on IP networks on that node.

Changing Persistent Node IP Labels

To change or view persistent node IP labels configured on a specified node:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Topology Configuration > Configure HACMP Persistent Node IP Labels/Addresses** > **Change/Show a Persistent Node IP Label/Address** and press Enter
3. Enter field values as follows:.

Node Name	The name of the node on which the IP Label/Address will be bound.
New Node Name	The new node name for binding the IP Label/Address.
Network Name	The name of the network on which the IP Label/Address will be bound.
Node IP Label/Address	The IP Label/Address to keep bound to the specified node.
New Node IP Label/Address	The new IP Label/Address to be bound to the specified node.

4. Press Enter. The resulting SMIT panel displays the current node name and persistent node IP labels defined on IP networks on that node.

Deleting Persistent Node IP Labels

To delete persistent node IP labels configured on a specified node,

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Topology Configuration > Configure HACMP Persistent Node IP Labels/Addresses** > **Remove a Persistent Node IP Label/Address**.
3. Press Enter.

   HACMP deletes the persistent node IP label from the node.

Changing the Configuration of a Global Network

Configuring a global network informs HACMP how different HACMP networks are connected to one another. This is commonly required when defining SP Ethernet networks that span subnets. You can group multiple HACMP networks *of the same type* under one logical global network name. This reduces the probability of network partitions that can cause the cluster nodes on one side of the partition to go down.

Networks combined into a global network cannot use IP Address Takeover (like the SP Ethernet).

The definition of a global network changes when you add or remove existing HACMP networks to or from the global network.

Adding an HACMP Network to a Global Network

To add a network to the global network definition:

1. Enter `smit hacmp`
2. Select **Extended Configuration > Extended Topology Configuration > Configure HACMP Global Networks** and press Enter.

   SMIT displays a picklist of defined HACMP networks.
3. Select an HACMP network and press Enter.

   SMIT displays the **Change/Show a Global Network** panel. The name of the network you selected is entered as the local network name.
4. Enter the name of the global network (character string) and press Enter.
5. Repeat these steps to define any new HACMP networks to be included in each global network.

Removing an HACMP Network from a Global Network

To remove a network from the global network definition, complete the following steps:

1. Enter `smit hacmp`
2. Select **Extended Configuration > Extended Topology Configuration > Configure Global Networks** and press Enter.

SMIT displays a picklist of defined HACMP networks.

3. Select the network to remove and press Enter.

   SMIT displays the **Change/Show a Global Network** panel. The name of the network you selected is entered as the local network name, along with the name of the global network where it currently belongs.

4. Remove the name of the global network and press Enter.
5. Repeat these steps to remove any other HACMP networks from a global network.

Changing the Configuration of a Network Module

The HACMP SMIT interface allows you to change the configuration of an HACMP network module. You may want to tune the parameters of the topology services by changing the failure detection rate of a network module.

This section contains the following topics:

Understanding Network Module Settings

The normal detection rate is usually optimal. Speeding up or slowing down failure detection rate is a small, but potentially significant area where you can adjust cluster fallover behavior. However, the amount and type of customization you add to event processing has a much greater impact on the total fallover time. You should test the system for some time before deciding to change the failure detection rate of any network module.

Be sure you have tuned the AIX 5L performance parameters for I/O pacing and **syncd** frequency before changing tuning parameters for a network module. See the section on Configuring AIX 5L for HACMP in the *Installation Guide.*

WARNING: I/O pacing and other tuning parameters should only be set to values other than defaults after a system performance analysis indicates that doing so will lead to both the desired and acceptable side effects. In addition, make sure you read the Setting I/O Pacing section in Chapter 1: Troubleshooting HACMP Clusters in the *Troubleshooting Guide* for more detailed description on tuning I/O pacing.

If you decide to change the failure detection rate of a network module, keep the following considerations in mind:

- Failure detection rate is dependent on the *fastest* network linking two nodes.

- Faster heartbeat rates may lead to false failure detections, particularly on busy networks. For example, bursts of high network traffic may delay heartbeats and this may result in nodes being falsely ejected from the cluster. Faster heartbeat rates also place a greater load on networks.
- If your networks are very busy and you experience false failure detections, you can change the failure detection rate on the network modules to **slow** to avoid this problem.

  For instance, in a mixed-version cluster with Token Ring networks, to allow enough time for any type of fallover to be handled properly by HACMP, you may want to adjust the Failure Detection Rate for this network module from **normal** to **slow.**

  Note: In rare cases, it is necessary to slow the Failure Detection Rate to even longer than the **slow** option SMIT offers. In this case, you may change the Failure Detection Rate of a network module to a custom value by changing the tuning parameters from their predefined values to custom values.

Resetting the Network Module Tunable Values to Defaults

For troubleshooting purposes, you or IBM support personnel assisting you with cluster administration may optionally reset the HACMP tunable values (such as the tuning parameters for the network module) to their installation-time defaults.

For more information on how to configure resetting the tunables in SMIT, see Resetting HACMP Tunable Values section in Chapter 1: Troubleshooting HACMP Clusters in the *Troubleshooting Guide*.

Behavior of Network Down on Serial Networks

Because of the point-to-point nature of serial networks, when there is any problem with the connection—such as the cable being unplugged—there is no possibility for other traffic to be visible on the other endpoint as there is with an IP network (like Ethernet). So when a serial interface loses heartbeats, it first declares its neighbor down, after the Failure Detection Rate has expired for that network interface type. HACMP waits the same interval again before declaring the local interface down (if no heartbeat is received from the neighbor).

Thus, the regular Failure Detection Rate formula applies to the detection of the remote interface down, and twice the Failure Detection Rate (Failure Cycle * Heartbeat Rate * 4) applies to the detection of the local interface down). HACMP does *not* run a **network_down** event until both the local and remote interfaces are failed. Therefore, for serial networks, the time from actual failure to the execution of the network down time is actually double the Failure Detection Rate value.

However, if the serial network is the last network left connecting this node to another, the **node_down** event is triggered by the first error.

In summary, for detecting a remote node down, the serial networks behave the same way as IP networks, and the time to detect a remote node down is still the longest Failure Detection Rate of the networks involved.

In addition, you can use the **clstat -s** command to display the service IP labels for serial networks that are currently down on a network.

Note: The RSCT **topsvcs** daemon logs messages whenever an interface changes state. These errors are visible in the **errpt**.

Disk Heartbeating Networks and Failure Detection

Disk heartbeating networks are identical to other non-IP based networks in terms of the operation of the failure detection rate, however there is a subtle difference that affects the state of the network endpoints and the events run:

- Disk heartbeating networks work by exchanging heartbeat messages on a reserved portion of a shared disk. As long as the node can access the disk the network endpoint will be considered up, even if heartbeat messages are *not* being sent between nodes. The disk heartbeating network itself will still be considered down.
- All other non-IP networks mark the network and both endpoints as down when either endpoint fails.

This difference makes it easier to diagnose problems with disk heartbeating networks: If the problem is in the connection of just one node with the shared disk only, then that part of the network will be marked as being down.

Disk Heartbeating and Fast Detection of Node Failures

HACMP 5.4 reduces the time it takes for a node failure to be realized throughout the cluster.

When a node fails, HACMP uses disk heartbeating to place a departing message on the shared disk so neighboring nodes are aware of the node failure within one heartbeat period (hbrate). Topology Services then distributes the information about the node failure throughout the cluster and each Topology Services daemon sends a **node_down** event to any concerned client.

You can turn on fast method for node failure detection when you configure disk heartbeating networks and specify a parameter for the disk heartbeating NIM.

For a procedure information, see the section Reducing the Node Failure Detection Rate: Enabling Fast Detection for Node Failures in this chapter.

For disk heartbeating information, see Configuring Heartbeating over Disk in Chapter 4: Configuring HACMP Cluster Topology and Resources (Extended).

Changing the Failure Detection Rate of a Network Module

Two parameters are involved in determining the Failure Detection Rate. They are:

- Heartbeat rate (in seconds)—frequency at which heartbeats (keepalives) are sent between the nodes.
- Failure detection cycle—the number of consecutive heartbeats that must be missed before failure is assumed.

The following two tables show the actual values of the Heartbeat Rate and the Failure Detection Cycle for IP and non-IP networks depending on predefined Failure Detection Rate (**Slow**, **Normal,** or **Fast**).

IP Network Setting	Seconds between Heartbeats	Failure Cycle	Failure Detection Rate
Slow	2	12	48
Normal	1	10	20
Fast	1	5	10

Failure Detection and Heartbeat Parameters for IP Networks

Non-IP Network Setting	Seconds between Heartbeats	Failure Cycle	Failure Detection Rate
Slow	3	8	48
Normal	2	5	20
Fast	1	5	10

Failure Detection and Heartbeat Parameters for Non-IP Networks

Before changing the default heartbeat settings for IP and non-IP networks, consult Chapter 3: Planning Cluster Network connectivity in the *Planning Guide* for information on and how these settings interact with the deadman switch.

Network Grace Period is the time period during IPAT via IP Replacement operations that node reachability is *not* computed for the network. The grace period value needs to be long enough for the network interface to be reconfigured with the new address and to rejoin its network interface membership group. When IPAT is used with HWAT, it usually takes longer for the operation to complete, so larger values of the grace period may be necessary. The default Grace Period value for Token Ring and ATM is 90 seconds. It is 60 seconds for all other network types.

SMIT provides two different panels for changing the attributes of a network module. You can either change the tuning parameters of a network module to predefined values of **Fast**, **Normal** and **Slow**, or you can set these attributes to custom values.

Changing the Tuning Parameters to Predefined Values

To change the tuning parameters of a network module to the predefined values of **Slow**, **Normal** or **Fast**:

1. Enter `smit hacmp`
2. Select **Extended Configuration > Extended Topology Configuration > Configure HACMP Network Modules** and press Enter.
3. Select the **Change a Network Module Using Predefined Values** option and press Enter. SMIT displays a list of defined network modules.

4. Select the network module you want to change and press Enter. SMIT displays the attributes of the network module, with their current values.

Network Module Name	Name of network type, for example, ether.
Description	For example, Ethernet Protocol
Failure Detection Rate	Select from **Normal, Fast** or **Slow**. This tunes the interval between heartbeats for the selected network module. The time needed to detect a failure can be calculated using this formula: (heartbeat rate) * (failure cycle) * 2 seconds.

5. Make the selections you need for your configuration.

   HACMP will detect a network interface failure in the time specified by the formula: Failure Detection Rate = Failure Cycle * Heartbeat Rate *2, or very close to it, the software may *not* take action on this event

   Due to event processing overhead the actual cluster event may *not* start for another few seconds.

6. Return to the SMIT **Extended Configuration** menu and select the **Extended Verification and Synchronization** option to synchronize the cluster.

Reducing the Node Failure Detection Rate: Enabling Fast Detection for Node Failures

Failure detection rates of **Fast**, **Normal** and **Slow** contain hbrates of 1, 2, or 3 seconds respectively. The time for the neighbor nodes to determine the node is down through disk heartbeating would be at most 1, 2, or 3 seconds, followed by the other cluster nodes becoming immediately aware of the failure.

Starting with HACMP 5.4, you can reduce the time it takes to detect a node failure. With the fast failure detection function, node failures are realized among the nodes in the cluster within one missed heartbeat period.

This method requires that you configure a disk heartbeating network. To enable this method, change the NIM parameter for the disk heartbeating network, when the cluster services are stopped on the nodes.

Fast failure detection method is supported on all disks that work with HACMP. It is not supported on SSA disks. For information on disk heartbeating, see Configuring Heartbeating over Disk.

To enable the fast method of detecting node failures:

1. Stop the cluster services on all nodes.
1. Enter the `SMIT hacmp`
2. In SMIT, go to **Extended Configuration > Extended Topology Configuration > Configure HACMP Network Modules > Change a Network Module using Custom Values** and press Enter. A list of network modules appears.
3. Select a network module that is used for the disk heartbeating network and press Enter.

4. Type or select values in entry fields as follows:

Network Module Name	Filled in with `diskhb`. This is the name of the network module for the disk heartbeating network.
Description	`Disk heartbeating network`
Address Type	`Device`. This specifies that the adapter associated with this network uses a device file.
Path	This is the path to the network executable file, such as `/usr/sbin/rsct/hats_diskhb_`
Parameters	This field lists the parameters passed to the network executable file. For the disk heartbeating NIM, enter FFD_ON in this field. This enables HACMP to use the fast method of node failure detection. This value in the NIM cannot be changed dynamically in a running cluster.

5. Leave the remaining fields in this SMIT screen unchanged for the disk heartbeating network.
6. Press Enter after making all desired changes and synchronize the cluster.

Changing the Tuning Parameters to Custom Values

If the cluster needs more customization than the predefined tuning parameters offer, you can change the Failure Detection Rate of a network module to a custom value. You can always return to the original settings by using the SMIT panel for setting the tuning parameters to predefined values.

Note: The failure detection rate of the network module affects the deadman switch time-out. The deadman switch time-out is triggered one second *before* the failure is detected on the *slowest* network in your cluster.

Also, use this SMIT panel to change the baud rate for TTYs if you are using RS232 networks that might *not* handle the default baud rate of 38400.

Setting Sub-Second Heartbeating Values

HACMP 5.2 and up lets you set sub-second heartbeating tunable values. These allow faster failure detection and therefore faster takeover operations.

This capability requires AIX 5L 5.2 or greater and RSCT 2.3.3 or greater on all cluster nodes.

Choose fast detection tunables with care, since the lower the detection time is, the greater the chance for false failures, that is, situations where a node or network interface is *not* really down, but appears to be because of a temporary problem.

For example, if the failure detection time for a network is set to 5 seconds, and the network or a node suffers a high load period or a period where packets are lost, then this may result in a node detecting that a remote node is down.

Before setting fast detection tunable values, take the following into account:

- The application load on the system should be such that it does *not* over-commit the amount of physical memory on the nodes. Having some amount of paging activity is acceptable, but the more paging activity exists on the system, the higher the probability that false failures may be seen because of processes being blocked while waiting for memory pages to be brought to memory.
- The rate of I/O interrupts on the node should *not* be such that processes in the system are prevented from getting timely access to the CPU.
- The traffic on the networks being used should be controlled, to avoid prolonged periods where cluster traffic cannot be reliably transmitted.

  In cases where the restrictions above cannot be followed, using low detection time values is *not* recommended.

To achieve a detection time of five seconds, use the following values:

NIM type	Failure Cycle	Interval Between Heartbeats (seconds)
All IP NIMs	5	0.5
RS232	3	0.8
Disk HB	3	0.8
TMSSA	3	0.8
TMSCSI	3	0.8

NIM Settings for 5 Second Detection Time

The failure detection time formula is: 2 x failure cycle x interval between heartbeats.

Still lower detection times may be used, but *not* with Disk HB and RS232 devices, since the minimum values for Failure Cycle and Interval Between Heartbeats for such devices are 3 and 0.75, respectively.

To achieve three seconds of detection time, use the following values:

NIM type	Failure Cycle	Interval between Heartbeats (seconds)
All IP NIMs	5	0.3
TMSSA	3	0.5
TMSCSI	3	0.5

NIM Settings for 3 Second Detection Time

Steps for Changing the Tuning Parameters of a Network Module to Custom Values

To change the tuning parameters of a Network Module to custom values:

1. Enter `smit hacmp`
2. Select **Extended Configuration > Extended Topology Configuration > Configure an HACMP Network Module > Change a Network Module Using Custom Values**.

SMIT displays a list of defined network modules.

3. Select the network module for which you want to change parameters and press Enter.

   SMIT displays the attributes of the network module, with their current values.

Network Module Name	Name of network type, for example, ether.
Description	For example, Ethernet Protocol
Address Type	Select an option: **Device** or **Address**. The **Address** option specifies that the network interface that is associated with this network module uses an IP-typed address. The **Device** option specifies that the network interface that is associated with this network module uses a device file.
Path	Specifies the path to the network executable file.
Parameters	Specifies the parameters passed to the network interface module (NIM) executable. For the RS232 NIM, this field specifies the baud rate. Allowable values are 38400 (the default), 19200, and 9600. For the disk heartbeating NIM, this field specifies the parameter that is passed to RSCT and that enables HACMP to use the fast method of node failure detection. Allowable values are FFD_ON and FFD_OFF (the default). To enable fast detection of node failures, specify FFD_ON in this field. This value in the NIM cannot be changed dynamically in a running cluster.
Grace Period	The current setting is the default for the network module selected. This is the time period in which, after a network failure was detected, further network failures of the same type would be ignored. This is 60 seconds for all networks except ATM and Token Ring, which are 90 seconds.
Failure Cycle	The current setting is the default for the network module selected. (Default for Ethernet is 10). This is the number of successive heartbeats that can be missed before the interface is considered to have failed. You can enter a number from 1 to 75.
Interval between Heartbeats (seconds)	The current setting is the default for the network module selected and is a heartbeat rate. This parameter tunes the interval (in seconds) between heartbeats for the selected network module. You can enter a number from less than 1 to 5.

Supports Gratuitous ARP This field is displayed *only* for those networks that generally support gratuitous ARP.

Set this field to **true** if this network supports gratuitous ARP. Setting this field to **true** enables HACMP to use IPAT via IP Aliases.

If you set this field to **false** for a specific network, you will disable the IPAT via IP Aliases function of HACMP for this network type. Since HACMP relies on the entry in this field to set up the fallover policies for cluster resources, *do not* change this field for the networks configured in your cluster.

Entry Type This field specifies the type of the network interface. It is either a network interface card (for a NIM specific to a network interface card), or a network interface type (for a NIM to use with a specific type of network device).

Next Generic Type This field specifies the next type of NIM to use if a more suitable NIM cannot be found.

Next Generic Name This field specifies the next generic NIM to use if a more suitable NIM cannot be found.

Supports Source Routing Set this field to **true** if this network supports IP loose source routing.

Note: Whenever a change is made to any of the values that affect the failure detection time—failure cycle (FC), heartbeat rate (HB) or failure detection rate—the new value of these parameters is sent as output to the panel in the following message:

```
SUCCESS: Adapter Failure Detection time is
now FC * HB* 2 or SS seconds
```

4. Make the changes for your configuration.

   HACMP will detect a network interface failure in the time specified by the formula: Failure Detection Rate = Failure Cycle * Heartbeat Rate *2, or very close to it, the software may *not* take action on this event

   Due to event processing overhead the actual cluster event may *not* start for another few seconds.

5. Synchronize the cluster.

After changing the `tty` baud rate and synchronizing the cluster, you can check the change by executing the following command on a running cluster (assuming `tty1` is the name of the HACMP heartbeat network):

```
stty -a < /dev/tty1
```

Showing a Network Module

To show the current values of a network module:

1. Enter `smit hacmp`
2. Select **Extended Configuration > Extended Topology Configuration > Configure an HACMP Network Module > Show a Network Module** and press Enter.

   SMIT displays a list of defined network modules.
3. Select the name of the network module for which you want to see current settings and press Enter.

   After the command completes, a panel appears that shows the current settings for the specified network module.

Removing a Network Module

To remove a network module:

1. Enter `smit hacmp`
2. Select **Extended Configuration > Extended Topology Configuration > Configure an HACMP Network Module > Remove a Network Module** and press Enter.

   SMIT displays a list of defined network modules.
3. Select the name of the network module you want to remove and press Enter. You will be asked `Are you sure?`
4. Press Enter again.

Changing an RS232 Network Module Baud Rate

All RS232 networks used by HACMP are brought up by RSCT with a default baud rate of 38400. However, there may be times when you need to lower that baud rate to a slower speed. To lower the baud rate for an already defined RS232 network, take these steps:

1. Enter `smit hacmp`
2. Select **Extended Configuration > Extended Topology Configuration > Configure an HACMP Network Module > Change a Network Module Using Custom Values**.
3. SMIT displays a list of Network Modules. Select RS232 and press Enter.
4. SMIT displays the **Change a Network Module using Custom Values** panel. Select the desired value in the **Parameters** field. 9600, 19200, or 38400 are the only acceptable baud rate values.
5. Press Enter.
6. Synchronize the cluster.

Changing the Configuration of a Site

If you are using one of the HACMP/XD solutions, be sure to consult the documentation for that software before changing any attributes of a site. Making changes to sites also affects cross-site LVM mirroring.

All dynamic topology configuration changes allowed in an HACMP configuration are now supported in HACMP/XD configurations. This includes changes to XD-type networks (XD_data used in HACMP/XD for GLVM), interfaces, sites, nodes, and NIM values. HACMP handles the resource group with replicated resources and its primary and secondary instances properly during these operations.

To avoid unnecessary processing of resources, use the Resource Group Migration utility, **clRGmove**, (**HACMP Resource Group and Application Management** in SMIT) to move resource groups that will be affected before you make the cluster topology change.

When dynamically reconfiguring a cluster, HACMP releases resource groups if this is found to be necessary, and then reacquires them later. For example, HACMP will release and reacquire the resource group that is using the associated service address on a network interface that is affected by the change to topology.

To change or show a site definition:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Topology Configuration > Configure HACMP Sites** > **Change/Show a Site**. and press Enter.
3. Select the site to change from the picklist.
4. Enter the information as follows:

Site Name	The current name is displayed.
New Site Name	Enter a name for this site using alphanumeric characters and underscores. Use no more than 32 characters.
Site Nodes	Add or remove names from the list of the cluster nodes that currently belong to the site.
Dominance	Select **yes** or **no** to indicate whether the current site is dominant or *not*. This only applies to HAGEO.
Backup Communications Type	Select the type of backup communication for your HAGEO cluster (**DBFS** for telephone line, **SGN** for a Geo_Secondary network, or **NONE**. HACMP/XD for Metro Mirror and HACMP/XD for GLVM only use **NONE**.

5. Press Enter to change the definition in the Configuration Database.

Note: If you change a site name that has an associated IP label, the IP label will change to associate with the new name for this site.

Removing a Site Definition

To remove a site definition:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Topology Configuration** > **Configure HACMP Sites** > **Remove a Site** and press Enter.
3. Select the site to remove from the picklist.

SMIT displays `Are you sure?`

4. Press Enter to remove the site definition.

Note: If you remove a site definition that has an associated IP label, the IP label remains, but is no longer associated with any site.

Synchronizing the Cluster Configuration

Whenever you modify the cluster definition in the Configuration Database on one node, you must synchronize the change with the Configuration Database data on all cluster nodes. You perform a synchronization by choosing the **Verification and Synchronization** option from either the Standard or the Extended HACMP Configuration SMIT panel, or from the **Problem Determination Tools** menu.

See Chapter 7: Verifying and Synchronizing an HACMP Cluster for complete information on this procedure.

Dynamic Reconfiguration Issues and Synchronization

This section is relevant for dynamic reconfiguration of both topology and resources.

Releasing a Dynamic Reconfiguration Lock

During a dynamic reconfiguration, HACMP creates a temporary copy of the HACMP-specific Configuration Database classes and stores them in the Staging Configuration Directory (SCD). This allows you to modify the cluster configuration while a dynamic reconfiguration is in progress. You cannot, however, synchronize the new configuration until the first is finished. The presence of an SCD on any cluster node prevents dynamic reconfiguration. If, because of a node failure or other reason, an SCD remains on a node after a dynamic reconfiguration is finished, it will prevent any further dynamic reconfiguration. Before you can perform further reconfiguration, you must remove this lock.

To remove a dynamic reconfiguration lock:

1. Enter `smit hacmp`
2. In SMIT, select **Problem Determination Tools** and press Enter.
3. Select the **Release Locks Set By Dynamic Reconfiguration** option and press Enter. SMIT displays a panel asking if you want to proceed. If you want to remove the SCD, press Enter.

Processing Configuration Database Data During Dynamic Reconfiguration

When you synchronize the cluster topology, the processing performed by HACMP varies depending on the status of the Cluster Manager.

The following describe the variations that may occur:

Cluster Manager Is Not Running on Any Cluster Node

If the Cluster Manager is *not* running on any cluster node (typically the case when a cluster is first configured), synchronizing the topology causes the configuration data stored on each node reachable from the local node to be updated.

Cluster Manager Is Running on the Local Node

If the Cluster Manager is running on the local node, synchronizing the topology triggers a dynamic reconfiguration event. While processing this event, HACMP updates the configuration data stored on each cluster node that is reachable. Further processing makes the new configuration the currently active configuration.

Cluster Manager Is Running on Some Cluster Nodes but Not on the Local Node

If the Cluster Manager is running on some cluster nodes but *not* on the local node, synchronizing the topology causes the configuration data stored on each node that is reachable from the local node to be updated. However, the processing performed during a dynamic reconfiguration to make the new configuration the active configuration is *not* performed.

Undoing a Dynamic Reconfiguration

Before HACMP overwrites the configuration defined in the ACD, it saves a record of the configuration in a cluster snapshot. Only the **.odm** portion of a cluster snapshot is created; the **.info** file is *not* created. (For more information about cluster snapshots, see Chapter 18: Saving and Restoring Cluster Configurations.) If you want to undo the dynamic reconfiguration, you can use this cluster snapshot to restore the previous configuration.

HACMP saves snapshots of the last ten configurations in the default cluster snapshot directory, **/usr/es/sbin/cluster/snapshots**, with the name **active.*x*.odm**, where *x* is a digit between 0 and 9, with 0 being the most recent.

Restoring the Configuration Database Data in the DCD

If a dynamic reconfiguration operation fails or is interrupted, you may want to restore the configuration in the DCD with the current active configuration, which is stored in the ACD. HACMP allows you to save in a snapshot the changes you made to the configuration in the DCD before you overwrite it.

To replace the Configuration Database data stored in the DCD with the Configuration Database data in the ACD, perform the following procedure.

1. Enter `smit hacmp`
2. In SMIT, select **Problem Determination Tools** and press Enter.
3. Select **Restore HACMP Configuration Database from Active Configuration** and press Enter.
4. Enter field values as follows:

Cluster Snapshot Name of System Default HACMP ODMs	In this field, specify the name you want assigned to the cluster snapshot HACMP creates before it overwrites the ODM data stored in the DCD with the ODM data from the ACD. You can use this snapshot to save the configuration changes you made.

Cluster Snapshot Description of System Default HACMP ODMs	Enter any text string you want stored at the beginning of the snapshot.

5. Press Enter. SMIT displays the results.

Chapter 14: Managing the Cluster Resources

This chapter describes how to manage the resources in your cluster. The first part of the chapter describes the dynamic reconfiguration process. The second part of the chapter describes procedures for making changes to individual cluster resources.

This chapter includes these topics:

- Dynamic Reconfiguration: Overview
- Changing or Removing Application Monitors
- Reconfiguring Service IP Labels as Resources in Resource Groups
- Reconfiguring Communication Links
- Reconfiguring Tape Drive Resources
- Using NFS with HACMP
- Reconfiguring Resources in Clusters with Dependent Resource Groups
- Synchronizing Cluster Resources.

For information on managing volume groups as resources, see Chapter 11: Managing Shared LVM Components and Chapter 12: Managing Shared LVM Components in a Concurrent Access Environment.

For information on setting up NFS properly, see Chapter 5: Planning Shared LVM Components in the *Planning Guide*.

Note: You can use either ASCII SMIT or WebSMIT to configure and manage the cluster. For more information on WebSMIT, see Chapter 2: Administering a Cluster Using WebSMIT.

Dynamic Reconfiguration: Overview

When you configure an HACMP cluster, configuration data is stored in HACMP-specific object classes in the ODM. The AIX 5L ODM object classes are stored in the default system configuration directory (DCD), **/etc/objrepos**.

You can make certain changes to both the cluster topology and to the cluster resources while the cluster is running (dynamic reconfiguration, or DARE). You can make a combination of resource and topology changes via one dynamic reconfiguration operation making the whole operation faster, especially for complex configuration changes.

Note: No automatic corrective actions take place during a DARE.

Reconfiguring a Cluster Dynamically

WARNING: Do *not* make configuration changes or perform any action that affects a resource if any node in the cluster has cluster services stopped and its resource groups placed in an UNMANAGED state.

At cluster startup, HACMP copies HACMP-specific ODM classes into a separate directory called the Active Configuration Directory (ACD). While a cluster is running, the HACMP daemons, scripts, and utilities reference the ODM data stored in the active configuration directory (ACD) in the ODM.

If you synchronize the cluster topology and cluster resources definition while the Cluster Manager is running on the local node, this action triggers a dynamic reconfiguration (DARE) event. In a dynamic reconfiguration event, the ODM data in the Default Configuration Directory (DCD) on all cluster nodes is updated and the ODM data in the ACD is overwritten with the new configuration data. The HACMP daemons are refreshed so that the new configuration becomes the currently active configuration.

The dynamic reconfiguration operation (that changes *both* resources and topology) progresses in the following order that ensures proper handling of resources:

- Releases any resources affected by the reconfiguration
- Reconfigures the topology
- Acquires and reacquires any resources affected by the reconfiguration operation.

For information on DARE in clusters with dependent resource groups, see Reconfiguring Resources in Clusters with Dependent Resource Groups.

Requirements before Reconfiguring

Before making changes to a cluster definition, ensure that:

- The same version of HACMP is installed on all nodes
- Some nodes are up and running HACMP and able to communicate with each other: *no node may have cluster services stopped with resource groups in an UNMANAGED state.* You must make changes form a node that is up so the cluster can be synchronized.
- The cluster is stable; the **hacmp.out** file does *not* contain recent event errors or **config_too_long** events.

Dynamic Cluster Resource Changes

DARE supports resource and topology changes done in one operation. Starting with HACMP 5.3, DARE is supported in HACMP/XD configurations.

You can make the following changes to cluster resources in an active cluster, dynamically:

- Add, remove, or change an application server.
- Add, remove, or change application monitoring.
- Add or remove the contents of one or more resource groups.
- Add, remove, or change a tape resource.
- Add or remove one or more resource groups.

- Add, remove, or change the order of participating nodes in a resource group.
- Change the node relationship of the resource group.
- Change resource group processing order.
- Add, remove or change the fallback timer policy associated with a resource group.

  The new value for the timer will come into effect after synchronizing the cluster and *after* the resource group is released and restarted (on a different node or on the same node) due to either a cluster event or if you move the group to another node.
- Add, remove or change the settling time for resource groups.
- Add or remove node distribution policy for resource groups.
- Add, change, or remove parent/child or location dependencies for resource groups. (Some limitations apply. See the section Making Dynamic Changes to Dependent Resource Groups.)
- Add, change, or remove inter-site management policy for resource groups.
- Add or remove a replicated resource. (You cannot change a replicated resource to non-replicated or vice versa.)
- Add, remove, or change pre- or post-events.

Additional Considerations with Dynamic Reconfiguration Changes

Depending on your configuration, you may need to take the following issues into account:

- **DARE changes to the settling time**. The current settling time continues to be active until the resource group moves to another node or goes offline. A DARE operation may result in the release and re-acquisition of a resource group, in which case the new settling time values take effect immediately.
- **Changing the *name* of an application server, node or resource group.** You must stop cluster services before they become active. You can include such a change in a dynamic reconfiguration; however, HACMP interprets these changes, particularly name change, as defining a new cluster component rather than as changing an existing component. Such a change causes HACMP to stop the active component before starting the new component, causing an interruption in service.
- **Adding, removing, or changing a resource group with replicated resources**. HACMP handles both the primary and secondary instances appropriately. For example, if you add a multi-site resource group to the configuration, HACMP will dynamically bring both primary and secondary instances online according to the site and node policies that you specify. You can also change a resource group's site management policy from **Ignore** to another option. HACMP then adds the secondary instances.

Dynamic reconfiguration is *not* supported during a cluster migration to a new version of HACMP, or when any node in the cluster has resource groups in the UNMANAGED state.

Reconfiguring Application Servers

An *application server* is a cluster resource used to control an application that must be kept highly available. It includes start and stop scripts.

Note that this section does *not* discuss how to write the start and stop scripts. See the vendor documentation for specific product information on starting and stopping a particular application.

Note: If you intend to add an application server dynamically, it is very important to test the server scripts beforehand, as they will take effect during the dynamic reconfiguration operation.

This section contains:

Changing an Application Server

When you specify new start or stop scripts to be associated with an application server, the HACMP configuration database is updated but the application server is *not* configured or unconfigured dynamically; thus the application controlled by the application server is *not* stopped and restarted. The next time the application is stopped, HACMP calls the *new* stop script—*not* the stop script that was defined when the application server was originally started.

Note: Changes to application server information are *not* automatically communicated to the application monitor configuration. Only the name of the server is updated on the SMIT panel for changing monitors. If you change an application server that has an application monitor defined, you must make the change *separately* to the application monitor as well. See Changing the Configuration of an Application Monitor.

To change (or view) an application server, complete the following steps:

1. Enter `smit hacmp`
2. In SMIT, select **Initialization and Standard Configuration > Configure Resources to Make Highly Available > Configure Application Servers** and press Enter. You can also use the **HACMP Extended Configuration** SMIT path to configure, change or remove application servers.
3. From this menu, select the **Change/Show an Application Server** option and press Enter. SMIT displays the application servers.
4. Select the application server to change and press Enter. The **Change/Show an Application Server** panel appears, with the server name filled in.

5. You can change the application name and/or the start and stop scripts.
6. Press Enter to add this information to the HACMP configuration database on the local node.
7. (*Optional*) Return to previous SMIT panels to perform other configuration tasks.
8. In SMIT, select the **Initialization and Standard Configuration** or **Extended Configuration** menu and select **Verification and Synchronization**. If the Cluster Manager is running on the local node, synchronizing the cluster resources triggers a dynamic reconfiguration event.

See Synchronizing Cluster Resources for more information.

Removing an Application Server

You can remove an application server from an active cluster dynamically. Before removing an application server, you must remove it from any resource group where it has been included as a resource. For more information, see Chapter 15: Managing Resource Groups in a Cluster.

Note: If you remove an application server, HACMP checks all application monitors for that server, and if only this server (and no other servers) use the associated monitors, it also removes the monitors. HACMP sends a message if monitors have been removed or preserved as a result of removing an application server.

To remove an application server, complete the following steps:

1. Enter `smit hacmp`
2. In SMIT, select **Initialization and Standard Configuration > Configure Resources to Make Highly Available > Configure Application Servers > Remove an Application Server** and press Enter.

   You can also use the **HACMP Extended Configuration** SMIT path to configure, change or remove application servers.

   SMIT displays the list of application servers.
3. Select the server to remove and press Enter. HACMP asks if you are sure you want to remove the server.
4. Press Enter again to confirm the removal. The server is removed from the HACMP configuration database on the local node.
5. (*Optional*) Return to previous SMIT panels to perform other configuration tasks.
6. Synchronize the cluster definition. If the Cluster Manager is running on the local node, synchronizing the cluster resources triggers a dynamic reconfiguration event.

See Synchronizing Cluster Resources for more information.

Changing or Removing Application Monitors

If you have configured application monitoring, you may wish to suspend or remove the monitor at some point. You can also change some aspect of the monitoring you have set up (for instance, the processes to be monitored, the scripts to run, or the notify, cleanup, or restart methods).

Note: This section discusses changing an existing application monitor. For information about adding a new application monitor, see Chapter 4: Configuring HACMP Cluster Topology and Resources (Extended).

Suspending and Resuming Application Monitoring

You can suspend the monitoring of a specified application while the cluster is running. This suspension of monitoring is temporary. If a cluster event occurs that results in the affected resource group moving to a different node, application monitoring resumes automatically on the new node. Similarly, if a node has resource group that are brought offline and then restarted, monitoring resumes automatically.

Note: If you have multiple monitors configured for one application, and if a monitor with **notify** action is launched first, HACMP runs the notification methods for that monitor, and the remaining monitor(s) are shut down on that node. HACMP takes no actions specified in any other monitor. You can restart the **fallover** monitor using the **Suspend/Resume Application Monitoring** SMIT panel.

To permanently stop monitoring of an application, do the steps in the section Removing an Application Monitor.

To temporarily suspend application monitoring:

1. Enter `smit hacmp`
2. In SMIT, select **HACMP System Management > Suspend/Resume Application Monitoring > Suspend Application Monitoring** and press Enter.

   You are prompted to select the application server for which this monitor is configured. If you have multiple application monitors, they are all suspended until you choose to resume them or until a cluster event occurs to resume them automatically, as explained above.

To resume monitoring after suspending it:

1. Enter `smit hacmp`
2. In SMIT, select **HACMP System Management > Suspend/Resume Application Monitoring > Resume Application Monitoring**.

   HACMP prompts you to select the application server that is associated with the suspended application monitor you want to resume.
3. Select the server. All monitors resume, configured as they were prior to suspending them.

Note: Do *not* make changes to the application monitor(s) configurations while they are in a suspended state.

Changing the Configuration of an Application Monitor

You can change the configuration details of an application monitor by editing the SMIT fields you defined when you configured the monitor initially.

Note: When you configured application monitors originally, the Restart Method and Cleanup Method fields had default values. If you changed those fields, and now want to change back to the defaults, you must enter the information manually (by copying the scripts from the **Change/Show an Application Server** SMIT panel).

To alter a defined application monitor, take the following steps.

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Resource Configuration** > **HACMP Extended Resources Configuration** > **Configure HACMP Applications > Configure HACMP Application Monitoring** and press Enter.
3. Depending on which type of monitor you are altering, select either:

   Define Process Application Monitor > Change/Show Process Application Monitor

   or

   Define Custom Application Monitor > Change/Show Custom Application Monitor.
4. From the list of monitors, select the previously defined application monitor you want to change.
5. Make changes in the SMIT panel fields and press Enter. Remember that default values are *not* restored automatically.

   The changes you enter take effect the next time the resource group containing the application is restarted.

Removing an Application Monitor

To permanently remove an application monitor:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Resource Configuration** > **HACMP Extended Resources Configuration** > **Configure HACMP Applications** > **Configure HACMP Application Monitoring** and press Enter.
3. Depending on which type of monitor you are altering, select either:

   Define Process Application Monitor > Remove a Process Application Monitor

   or

   Define Custom Application Monitor > Remove a Custom Application Monitor.
4. Select the monitor to remove.
5. Press Enter. The selected monitor is deleted.

If the monitor is currently running, it is *not* stopped until the next dynamic reconfiguration or synchronization occurs.

Note: If you remove an application monitor, HACMP removes it from the server definition for all application servers that were using the monitor, and sends a message about the servers that will no longer use the monitor.

If you remove an application server, HACMP removes that server from the definition of all application monitors that were configured to monitor the application. HACMP also sends a message about which monitor will no longer be used for the application. If you remove the last application server in use for any particular monitor; that is, if the monitor will no longer be used for any application, verification issues a warning that the monitor will no longer be used.

Reconfiguring Service IP Labels as Resources in Resource Groups

This section contains:

- Steps for Changing the Service IP Labels/Addresses Definitions
- Deleting Service IP Labels
- Changing Distribution Preference for Service IP Label Aliases
- Viewing Distribution Preference for Service IP Label Aliases.

You must stop cluster services to change service IP labels/address resources that are already included in resource group.

Remember to add any new service IP labels/addresses to the **/etc/hosts** file before using them. If you intend to change the names of existing labels, first create the new names and add them to the **etc/hosts** file. Then make the name change in SMIT.

Do *not* remove the previously used service IP label/address from the **/etc/hosts** file until after you have made the change in the cluster configuration. Once you make the change in the configuration and in the **/etc/hosts** file on the local node, make the change in the **/etc/hosts** files of the other nodes before you synchronize and restart the cluster.

Steps for Changing the Service IP Labels/Addresses Definitions

To change a service IP label/address definition, take the following steps:

1. Stop cluster services on all nodes.
2. On any cluster node, enter `smit hacmp`
3. Select **HACMP Initialization and Standard Configuration** > **Configure Resources to Make Highly Available > Configure Service IP Labels/Addresses > Change/Show a Service IP Label/Address**.

   Note: In the Extended Cluster Configuration flow, the SMIT path is **HACMP > Extended Configuration > HACMP Extended Resources Configuration > Configure Service IP Labels/Addresses > Change/Show a Service IP Label/Address**.

4. In the **IP Label/Address to Change** panel, select the IP Label/Address you want to change. The **Change/Show a Service IP Label/Address** panel appears.
5. Make changes in the field values as needed.

6. Press Enter after filling in all required fields. HACMP now checks the validity of the new configuration. You may receive warnings if a node cannot be reached, or if network interfaces are found to *not* actually be on the same physical network.
7. On the local node, verify and synchronize the cluster.

   Return to the HACMP Standard or Extended Configuration SMIT panel and select the **Verification and Synchronization** option.
8. Restart Cluster Services.

Deleting Service IP Labels

To delete an IP Label/Address definition, take the following steps:

1. Stop cluster services on all nodes.
2. On any cluster node, enter `smit hacmp`
3. Select **Initialization and Standard Configuration** > **Configure Resources to Make Highly Available > Configure Service IP Labels/Addresses > Delete a Service IP Label/Address**. A panel appears with the list of IP labels/addresses configured to HACMP.

   Note: In the Extended Cluster Configuration flow, the SMIT path is **Extended Configuration > Extended Resources Configuration > Configure Service IP Labels/Addresses > Delete a Service IP Label/Address**.
4. Select one or more labels that you want to delete from the list and press Enter.
5. HACMP displays `Are You Sure?` If you press Enter, the selected labels/addresses are deleted.
6. For maintenance purposes, delete the labels/addresses from the **/etc/hosts** file.

   After you delete service IP labels from the cluster configuration using SMIT, removing them from **/etc/hosts** is a good practice because it reduces the possibility of having conflicting entries if the labels are reused with different addresses in a future configuration.

Changing AIX 5L Network Interface Names

When you define communication interfaces by entering or selecting an HACMP IP label/address, HACMP discovers the associated AIX 5L network interface name. HACMP expects this relationship to remain unchanged. If you change the name of the AIX 5L network interface name after configuring and synchronizing the cluster, HACMP will *not* function correctly.

If this problem occurs, you can reset the communication interface name from the SMIT HACMP **Cluster System Management (C-SPOC**) menu.

To reset the HACMP communication interface:

1. Enter `smit hacmp`
2. In SMIT, select **Cluster System Management (C-SPOC)** > **HACMP Communications Interface Management** > **Update HACMP Communication Interface with AIX Settings** and press Enter.

3. Press F4 and select the communication interface that you want to reset from the HACMP picklist.
4. Press Enter to complete the reset operation.
5. On the local node, verify and synchronize the cluster. Return to the **Extended** or **Standard Configuration** SMIT panel and select the **Verification and Synchronization** option.

See Synchronizing Cluster Resources for more information.

Changing Distribution Preference for Service IP Label Aliases

You can configure a distribution preference for the service IP labels that are placed under HACMP control. HACMP lets you specify the distribution preference for the service IP label aliases. These are the service IP labels that are part of HACMP resource groups and that belong to IPAT via IP Aliasing networks.

When you specify the new distribution preference to be associated with a network, the HACMP configuration database is updated but the preference is *not* changed dynamically; that is, HACMP does *not* interrupt the processing by relocating service IP labels at the time the preference is changed. Instead, the next time a cluster event, such as a fallover takes place for a resource group that has service IP labels on the network, HACMP uses the new distribution preference when it allocates the service IP label alias on the network interface on the backup node.

For information on types of distribution preferences, see Types of Distribution for Service IP Label Aliases in Chapter 4: Configuring HACMP Cluster Topology and Resources (Extended).

To change a defined distribution preference for service IP labels:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Resource Configuration > HACMP Extended Resources Configuration > Configure Resource Distribution Preferences > Configure Service IP labels/addresses Distribution Preferences** and press Enter.

   HACMP displays a list of networks that use IPAT via IP Aliases.
3. From the list of networks, select the network for which you want to change the distribution preference and press Enter.
4. Change the distribution preference and press Enter. Remember that default values are *not* restored automatically.

   The changes you enter take effect the next time the resource group containing the service IP label is restarted.

Viewing Distribution Preference for Service IP Label Aliases

Use the **cltopinfo** command to display the service IP label distribution preference specified for a particular network.

For information on this command, see the *Troubleshooting Guide.*

Alternatively, you can use the **cllsnw - c** command to display the service IP label distribution preference (`sldp`) specified for a particular network.

The syntax is as follows:

```
cllsnw - c
#netname:attr:alias:monitor_method:sldp:
```

Where `sldp` stands for the service label distribution preference.

Example:

```
net_ether_01:public:true:default:ppstest::sldp_collocation_with_persistent
```

Reconfiguring Communication Links

Highly available communication links can be of three types: SNA-over-LAN, X.25, or SNA-over-X.25.

Changes to a communication link may involve changing the adapter information or changing link information such as the name, the ports or link stations, or the application server (start and stop) scripts. You can reconfigure communication links using the SMIT interface.

To change the configuration of a highly available communication link, you may need to change both the adapter information and the link information.

Note: When a resource group has a list of Service IP labels and Highly Available Communication Links with configured SNA resources, the first Service IP label in the list of Service IP labels defined in the resource group will be used to configure SNA.

This section contains:

- Changing Communication Adapter Information
- Removing a Communication Adapter from HACMP
- Changing Communication Link Information
- Removing a Communication Link from HACMP.

Changing Communication Adapter Information

To change or view the configuration information for an X.25 communication adapter, complete the following steps:

1. Enter `smit hacmp`
2. In SMIT, select the **Extended Configuration** > **Extended Cluster Resources > HACMP Extended Resources Configuration > Configure Communication Adapters and Links for HACMP > Configure Communication Adapters for HACMP > Change/Show Communications Adapter** and press Enter.
3. Select the adapter to change and press Enter.
4. Make the changes as needed. To review the instructions for the field entries, refer to Configuring Highly Available Communication Links in Chapter 4: Configuring HACMP Cluster Topology and Resources (Extended).

5. Press Enter to add this information to the Configuration Database on the local node.
6. Return to previous SMIT panels to perform other configuration tasks.
7. To verify and synchronize the changes, return to the **Extended Configuration** SMIT panel and select the **Verification and Synchronization** option.

If the Cluster Manager is running on the local node, synchronizing the cluster triggers a dynamic reconfiguration event. See Synchronizing Cluster Resources for more information.

Removing a Communication Adapter from HACMP

To remove a communication adapter, complete the following steps:

1. Enter `smit hacmp`
2. In SMIT, select the **Extended Configuration > Extended Cluster Resources > HACMP Extended Resources Configuration > Configure Communication Adapters and Links for HACMP > Configure Communication Adapters for HACMP > Remove a Communications Adapter** and press Enter.
3. Select the adapter to remove and press Enter. A message asks if you are sure you want to remove the communication adapter.
4. Press Enter again to confirm the removal. The adapter is removed from the Configuration Database object classes stored in the DCD on the local node.
5. Return to previous SMIT panels to perform other configuration tasks.
6. To verify and synchronize the changes, return to the **Extended** or **Standard Configuration** SMIT panel and select the **Verification and Synchronization** option.

If the Cluster Manager is running on the local node, synchronizing the cluster triggers a dynamic reconfiguration event. See Synchronizing Cluster Resources for more information.

Changing Communication Link Information

To change or view a communication link, complete the following steps:

1. Enter `smit hacmp`
2. In SMIT, select the **Extended Configuration > Extended Cluster Resources > HACMP Extended Resources Configuration > Configure Communication Adapters and Links for HACMP > Configure Highly Available Communication Links > Change/Show Highly Available Communication Link** and press Enter.

   SMIT displays a list of links.
3. Select the link to change and press Enter.
4. Make the changes as needed. To review the instructions for the field entries, refer to Configuring Highly Available Communication Links in Chapter 4: Configuring HACMP Cluster Topology and Resources (Extended).
5. Press Enter to add this information to the Configuration Database stored in the DCD on the local node.
6. When all changes have been made, return to previous SMIT panels to perform other configuration tasks.

7. To verify and synchronize the changes, return to the **Extended Configuration** SMIT panel and select the **Verification and Synchronization** option.

If the Cluster Manager is running on the local node, synchronizing the cluster triggers a dynamic reconfiguration event. See Synchronizing Cluster Resources for more information.

Removing a Communication Link from HACMP

You can remove a communication link from an active cluster dynamically. Before removing a communication link, you must remove it from any resource group where it is included as a resource.

To remove a communication link, complete the following steps:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Resources Configuration > HACMP Extended Resource Group Configuration > Change/Show Resources/Attributes for a Resource Group** and press Enter.

   SMIT displays a list of resource groups.
3. Select the appropriate resource group, and in the **Communication Links** field, remove the link(s) from the list.
4. Next, remove the link definition from HACMP. In SMIT, select the **Extended Resources Configuration > HACMP Extended Resources Configuration > Configure Communication Adapters and Links for HACMP > Configure Highly Available Communication Links > Remove Highly Available Communication Link** and press Enter.

   SMIT displays a list of links.
5. Select the communication link you want to remove and press Enter. A message asks if you are sure you want to remove the communication link.
6. Press Enter again to confirm the removal. The link is removed from the Configuration Database on the local node.
7. Return to previous SMIT panels to perform other configuration tasks.
8. To verify and synchronize the changes, return to the **Extended Configuration** SMIT panel and select the **Verification and Synchronization** option.

If the Cluster Manager is running on the local node, synchronizing the cluster triggers a dynamic reconfiguration event. See Synchronizing Cluster Resources for more information.

Reconfiguring Tape Drive Resources

Using HACMP SMIT panels you can take the following actions to reconfigure tape drives:

- Add tape drives as HACMP resources
  - Specify synchronous or asynchronous tape operations
  - Specify appropriate error recovery procedures
- Change/Show tape drive resources

- Remove tape drive resources
- Add or remove tape drives to/from HACMP resource groups.

To add tape drive resources, see Chapter 3: Configuring an HACMP Cluster (Standard).

Changing a Tape Resource

To change or show the current configuration of a tape drive resource, take the following steps:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Resource Configuration > HACMP Extended Resources Configuration > Configure HACMP Tape Resources** > **Change/Show a Tape Resource** and press Enter.

   SMIT returns a picklist of the configured tape drive resources.
3. Select the tape resource you want to see or change.

   SMIT displays the current configuration for the chosen tape device.
4. Change the field values as necessary.
5. Press Enter.

Removing a Tape Device Resource

To remove a tape device resource, take the following steps:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration >Extended Resource Configuration > HACMP Extended Resources Configuration > Configure HACMP Tape Resources** > **Remove a Tape Resource** and press Enter.

   SMIT returns a picklist of the configured tape drive resources.
3. Select the tape resource you want to remove.

   SMIT displays the message `Are You Sure?`

Using NFS with HACMP

HACMP includes the following functionality:

- Reliable NFS server capability that allows a backup processor to recover current NFS activity should the primary NFS server fail, preserving the locks on NFS filesystems and dupcache. *This functionality is available for two-node clusters only*.
- Ability to specify a network for NFS mounting.
- Ability to define NFS exports and mounts at the directory level.
- Ability to specify export options for NFS-exported directories and filesystems.

See the section on Using NFS with HACMP in Chapter 5: Planning Shared LVM Components in the *Planning Guide* for complete information on this subject, including an example of NFS cross-mounting.

Reconfiguring Resources in Clusters with Dependent Resource Groups

If you have configured dependent resources in the cluster, the dynamic reconfiguration (DARE) lets you:

- Make changes to the cluster resources
- Make changes to the cluster topology
- Dynamically add or remove resource groups from the cluster configuration.

When reconfiguring resources dynamically, HACMP ensures the availability of applications in resource groups. For resource groups that have dependencies between them, it means that HACMP only allows changing resources when it is safe to do so.

This section describes the conditions under which HACMP performs dynamic reconfigurations in clusters with dependent resource groups.

Reconfiguring Resources and Topology Dynamically

Consider a cluster where resource group A (child) depends on resource group B (parent). In turn, resource group B depends on resource group C. Note that resource group B serves both as a parent for resource group A and a child for resource group C.

The following rules for DARE apply:

- You can make changes to the cluster topology and cluster resources dynamically for a child resource group and for a parent resource group.
- For a child resource group, if this resource group has no other groups that depend on it, HACMP runs the reconfiguration events and performs the requested changes. HACMP performs a dynamic reconfiguration of a child resource group without taking any other resource groups offline and online.
- For a parent resource group, before proceeding with dynamic reconfiguration events, you must manually take offline *all* child resource groups that depend on the parent resource group. After the dynamic reconfiguration is complete, you can bring the child resource groups back online.

  For instance, in a A>B>C dependency, where A is a child resource group that depends on B, and B is a child resource group that depends on C, to make changes to the resource group C, you must first take offline resource group A, then resource group B, and then perform a dynamic reconfiguration for resource group C. Once HACMP completes the event, you can bring online resource group B and then resource group A.

  If you attempt a dynamic reconfiguration event and HACMP detects that the resource group has dependent resource groups, the DARE operation fails and HACMP displays a message prompting you to take the child resource groups offline, before attempting to dynamically change resources or make topology changes in the parent resource group.

Making Dynamic Changes to Dependent Resource Groups

If you have dependent resource groups configured, the following rules apply:

- If you dynamically add a resource group to the cluster, HACMP processes this event without taking any resource groups offline or online.
- If you dynamically remove a resource group from the cluster configuration and the resource group is included in a dependency with one or more resource groups, then:
  - If a resource group that you remove dynamically is a parent resource group, then before processing the dynamic reconfiguration event to remove the group, HACMP temporarily takes offline dependent (child) resource group(s). After the DARE event is complete, HACMP reacquires child resource groups.

    For instance, consider the following resource group dependency: A >B>C, where A (child) depends on B, and B depends on C (parent). B is a child to resource group C and is a parent to resource group A.

  In this case, if you dynamically remove resource group C from the cluster configuration, HACMP takes resource group A offline, then it takes resource group B offline, removes resource group C, and reacquires first resource group B and then resource group A.

Cluster Processing During DARE in Clusters with Dependent Resource Groups

As with cluster processing for other events, if you have dependencies or sites configured in the cluster, cluster processing for dynamic reconfiguration is done in a different way than in clusters without dependencies between resource groups. As a result, the sequence of events in the **hacmp.out** file shows a series of **rg_move** events.

See the Job Types: Processing in Clusters with Dependent Resource Groups section in Chapter 2: Using Cluster Log Files in the *Troubleshooting Guide*, for information on how to interpret events that HACMP runs in clusters with dependencies.

Synchronizing Cluster Resources

Whenever you modify the configuration of cluster resources in the Configuration Database on one node, you must synchronize the change across all cluster nodes. You perform a synchronization by choosing the **Verification and Synchronization** option from either the Standard or the Extended HACMP configuration SMIT panel.

Note: If the cluster is running, make sure no node has been stopped with its resource groups placed in UNMANAGED state when performing a synchronization.

The processing performed in synchronization varies depending on whether the Cluster Manager is active on the local node:

- If the Cluster Manager is *not* active on the local node when you select this option, the Configuration Database data in the DCD on the local node is copied to the Configuration Databases stored in the DCDs on all cluster nodes.

- If the Cluster Manager is active on the local node, synchronization triggers a cluster-wide, dynamic reconfiguration event. In dynamic reconfiguration, the configuration data stored in the DCD is updated on each cluster node and, in addition, the new Configuration Database data replaces the Configuration Database data stored in the ACD on each cluster node. The cluster daemons are refreshed and the new configuration becomes the active configuration. In the HACMP log file, **reconfig_resource_release**, **reconfig_resource_acquire**, and **reconfig_resource_complete** events mark the progress of the dynamic reconfiguration.

See Chapter 7: Verifying and Synchronizing an HACMP Cluster for complete information on the SMIT options.

Notes

In some cases, the verification uncovers errors that do *not* cause the synchronization to fail. HACMP reports the errors in the SMIT command status window so that you are aware of an area of the configuration that may be a problem. You should investigate any error reports, even when they do *not* interfere with the synchronization.

Log files that are no longer stored in a default directory, but a user-specified directory instead, are verified by the cluster verification utility, which checks that each log file has the same pathname on every node in the cluster and reports an error if this is *not* the case.

Chapter 15: Managing Resource Groups in a Cluster

This chapter describes how to reconfigure the cluster resource groups. It describes adding and removing resource groups, and changing resource group attributes and processing order. It also covers the *Resource Group Management* utility that allows you to change the status and location of resource groups dynamically using the SMIT interface, or the **clRGmove** command. This utility lets you move resource groups to other cluster nodes, for instance to perform system maintenance on a particular cluster node.

If you have dependent resource groups in the cluster, see the section Reconfiguring Resources in Clusters with Dependent Resource Groups in Chapter 14: Managing the Cluster Resources for information on making dynamic reconfiguration changes to the cluster resources.

Changes to the cluster resource groups are grouped under two general categories in this chapter:

- Changes to Resource Groups
- Resource Group Migration.

Note: You can use either ASCII SMIT or WebSMIT to configure and manage the cluster and view interactive cluster status. You can also use WebSMIT to navigate, configure and view the status of the running cluster and graphical displays of sites, networks, nodes and resource group dependencies.

You can*not* migrate resource groups with replicated resources using WebSMIT.

Changes to Resource Groups

Changes you make to resource groups consist of the following actions:

- Reconfiguring Cluster Resources and Resource Groups
- Adding a Resource Group
- Removing a Resource Group
- Changing Resource Group Processing Order
- Resource Group Ordering during DARE
- Changing the Configuration of a Resource Group
- Changing Resource Group Attributes
- Changing a Dynamic Node Priority Policy
- Changing a Delayed Fallback Timer Policy
- Showing, Changing, or Deleting a Settling Time Policy

- Changing a Location Dependency between Resource Groups
- Changing a Parent/Child Dependency between Resource Groups
- Displaying a Parent/Child Dependency between Resource Groups
- Removing a Dependency between Resource Groups
- Adding or Removing Individual Resources
- Reconfiguring Resources in a Resource Group
- Forcing a Varyon of a Volume Group.

Reconfiguring Cluster Resources and Resource Groups

When you initially configured your HACMP system, you defined each resource as part of a resource group. This allows you to combine related resources into a single logical entity for easier configuration and management. You then configured each resource group to have a particular kind of relationship with a set of nodes. You also assigned a *priority* to each participating node for some non-concurrent resource groups.

To change the nodes associated with a given resource group or to change the priorities assigned to the nodes in a resource group chain, you must redefine the resource group. You must also redefine the resource group if you add or change a resource assigned to the group.

You can also redefine the order in which HACMP attempts to acquire and release the resource groups in your cluster. In general, HACMP processes all individual resource groups configured in your cluster in parallel unless you define a specific serial order upon which certain resource groups should be acquired or released, using the **Change/Show Resource Group Processing Order** panel in SMIT.

For general information about customizing the serial order of processing resource groups, see Chapter 6 in the *Planning Guide*.

This section describes how to view, change, add, and delete a resource group. For more information about the initial configuration of cluster resources, see Chapter 6 in the *Planning Guide* and the chapters in this Guide that describe initial configuration. These include:

- Chapter 3: Configuring an HACMP Cluster (Standard)
- Chapter 4: Configuring HACMP Cluster Topology and Resources (Extended)
- Chapter 5: Configuring HACMP Resource Groups (Extended).

Adding a Resource Group

You can add a resource group to an active cluster. You do *not* need to stop and then restart cluster services for the resource group to become part of the current cluster configuration.

To add a resource group, see Chapter 5: Configuring HACMP Resource Groups (Extended).

If the Cluster Manager is running on the local node, synchronizing the cluster triggers a dynamic reconfiguration event. For more information, see Chapter 7: Verifying and Synchronizing an HACMP Cluster.

Removing a Resource Group

You can remove a resource group from an active cluster. You do *not* need to stop and then restart cluster services for the resource group to be removed from the current cluster configuration.

To remove a resource group:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Resource Configuration > Extended Resource Group Configuration > Remove a Resource Group** and press Enter.

   SMIT displays a panel listing the defined resource groups.
3. Select the resource group you want to remove and press Enter. SMIT displays a popup warning, reminding you that all information about the resource group will be lost.

   Note: If you have the following parent/child resource group dependency chain configured: A > B > C, and remove the resource group B, HACMP sends a warning that the dependency links between A and B, and between B and C are also removed. For more information, see Configuring Dependencies between Resource Groups in Chapter 5: Configuring HACMP Resource Groups (Extended).
4. Press Enter again to confirm your action.
5. Return to previous SMIT panels to perform other configuration tasks.
6. To synchronize the cluster definition, In SMIT, select the **Extended Configuration > Extended Verification and Synchronization** SMIT panel and press Enter.

   If the Cluster Manager is running on the local node, synchronizing cluster resources triggers a dynamic reconfiguration event. For more information, see Chapter 7: Verifying and Synchronizing an HACMP Cluster.

Changing Resource Group Processing Order

By default, HACMP acquires and releases resource groups in parallel. To view or change the current order in which HACMP processes resource groups in your cluster use the **Change/Show Resource Group Processing Order** panel in SMIT. See Configuring Processing Order for Resource Groups in Chapter 5: Configuring HACMP Resource Groups (Extended).

Resource Group Ordering during DARE

In general, HACMP processes all individual resource groups configured in your cluster in parallel unless you define a specific serial order upon which certain resource groups should be acquired or released. Handling of any dependencies between resource groups take priority over any serial processing you specify.

If you need to control the actual processing order during dynamic reconfiguration (DARE), make the changes to only one resource group at a time. Otherwise, the order in which resource groups are acquired and released may be unpredictable.

During the dynamic reconfiguration process, you could have two scenarios:

- Prior to dynamically changing any of the resource groups:
  - The processing order for *all* the resource groups was parallel

  and

  - You did *not* change it during dynamic reconfiguration (DARE).

  In this case, during the dynamic reconfiguration process, HACMP processes the resource groups according to an alphabetically-sorted order, and *not* in parallel. If you made the changes to particular resource groups in the cluster, these changes may affect the order in which these resources will be actually released and acquired.
- Prior to dynamically changing any of the resource groups:
  - The processing order for some of the resource groups was parallel

    and

  - Some of the resource groups were included in the list for serial processing.

  In this case, if during DARE, you change the serial order in which some of the resource groups are acquired or released on the nodes, then the newly specified order becomes valid during the reconfiguration process. The new order is used by HACMP during the same cluster reconfiguration cycle.

After reconfiguration is complete, HACMP returns to the usual order of processing, as described below.

Resource group acquisition in HACMP occurs in the following order:

1. Resource groups for which the customized order is specified are acquired in the customized serial order.
2. If some of the resource groups in the cluster have dependencies between them, these resource groups are processed in phases. For example, parent resource groups are acquired before child resource groups are acquired.
3. Resource groups that mount NFS only are processed in the specified order.
4. Resource groups that are *not* included in the customized ordering lists are acquired in parallel.

Resource group release in HACMP occurs in the following order:

1. Resource groups for which no customized order have been specified are released in parallel.
2. HACMP releases resource groups that are included in the customized release ordering list.
3. If some of the resource groups in the cluster have dependencies between them, these resource groups are processed in phases. For example, the child resource groups are released before the parent resource groups are released.
4. Resource groups that must unmount NFS are processed in the specified order.

However, if you made changes to particular resource groups in the cluster, these changes may affect the order in which these resource groups are released and acquired. As a result, during the dynamic reconfiguration process, the actual order in which resource groups are acquired and released is unpredictable.

This order is dependent on the changes you make to the order during DARE, and on the types of dynamic changes you make to the resource groups themselves. For instance, due to the changes you made to a particular resource group, this resource group may need to be released before others in the list, even though the alphabetically-sorted order is used for the remaining resource groups.

Changing the Configuration of a Resource Group

You can change the following for a configured resource group:

- The name of the resource group
- The nodes in the list of participating nodes
- The site management policy of a resource group
- The priority of participating nodes (by changing their position in the list of participating nodes)
- The startup, fallover and fallback policies for resource groups
- Attributes of the resource group.

Warning: If you have added resources to the resource group, you need to remove them prior to changing the resource group's startup, fallover and fallback policies.

You can change most of the attributes of a resource group in an active cluster without having to stop and then restart cluster services. However, to change the name of a resource group, you must stop and then restart the cluster to make the change part of the current cluster configuration.

Changing the Basic Configuration of a Resource Group

To change the basic configuration of a resource group:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Resource Configuration > HACMP Extended Resource Group Configuration > Change/Show a Resource Group**. SMIT displays a list of the currently defined resource groups.
3. Select the resource group to change and press Enter.

   Note: HACMP shows *only* the valid choices for the specified resource group.

4. Enter field values as necessary.
5. Press Enter to change the resource group information stored in the HACMP Configuration Database (ODM).
6. Return to previous SMIT panels to perform other configuration tasks or to synchronize the changes you just made.

   If the Cluster Manager is running on the local node, synchronizing cluster resources triggers a dynamic reconfiguration event. For more information, see Chapter 7: Verifying and Synchronizing an HACMP Cluster.

Changing Resource Group Attributes

To change the attributes of a resource group:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Resource Configuration > HACMP Extended Resource Group Configuration > Change/Show Resources/Attributes for a Resource Group**. SMIT displays a list of the currently defined resource groups.
3. Select the resource group you want to change and press Enter.

   SMIT displays a list of resource group attributes and the values set.
4. Change field values as needed.
5. Press Enter to change the resource group information stored in the HACMP Configuration Database.
6. Return to previous SMIT panels to perform other configuration tasks.
7. To synchronize the changes you made, In SMIT, select the **Extended Configuration** SMIT panel and select the **Extended Verification and Synchronization** option.

   If the Cluster Manager is running on the local node, synchronizing cluster resources triggers a dynamic reconfiguration event.

Changing a Dynamic Node Priority Policy

You can use SMIT to change or show a dynamic node priority policy.

To show or change the dynamic node priority policy for a resource group:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Resource Configuration > Extended Resource Group Configuration > Change/Show Resources /Attributes for a Resource Group** and press Enter.
3. Select the resource group.

   You can change the dynamic node priority policy on the next panel, if you have one configured previously.
4. Select the policy you want and press Enter. See Dynamic Node Priority Policies in Chapter 5: Configuring HACMP Resource Groups (Extended) for more information.

Changing a Delayed Fallback Timer Policy

You can use SMIT to change or show a delayed fallback timer policy.

To change or show a previously configured fallback policy, follow these steps:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration** > **Extended Resource Configuration > Extended Resource Group Configuration > Configure Resource Group Run-Time Policies > Change/Show a Delayed Fallback Timer Policy** and press Enter. A picklist displays the previously configured timer policies.

3. Select the fallback timer policy to change.
4. Change the fallback timer policy on the next panel.

   The new value for the timer will come into effect after synchronizing the cluster and *after* the resource group is released and restarted (on a different node or on the same node) due to either a cluster event or if you move the group to another node.

   Note that you can change the parameters, but you cannot change the type of recurrence for the specific fallback timer. However, you can configure another fallback timer policy that uses a different predefined recurrence, and assign it to a resource group.

Removing a Delayed Fallback Timer Policy for a Resource Group

Note that you cannot remove a delayed fallback timer if any resource groups are configured to use it. First, change or remove the delayed fallback timer included as an attribute to any resource groups configured to use the unwanted timer, then proceed to remove it, as described in the following procedure.

To delete a previously configured delayed fallback timer policy:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration> Extended Resource Configuration > Extended Resource Group Configuration > Configure Resource Group Run-Time Policies > Delete a Delayed Fallback Timer Policy** and press Enter. A picklist displays the previously configured timer policies.
3. Select the fallback timer policy to remove and press Enter. You will be asked `Are you sure?`
4. Press Enter again.

Showing, Changing, or Deleting a Settling Time Policy

You can change, show or delete previously configured settling time policies using the **Extended HACMP Resource Group Configuration > Configure Resource Group Run-Time Policies > Configure Settling Time Policy** SMIT path.

Changing a Location Dependency between Resource Groups

There are three types of location dependencies between resource groups:

- Online On Same Node
- Online On Different Nodes
- Online On Same Site

Changing an Online on Same Node Dependency

To change an **Online on Same Node** location dependency between resource groups:

1. Enter `smit hacmp`

2. In SMIT, select **HACMP Extended Resource Configuration > Configure Resource Group Run-Time Policies > Configure Dependencies between Resource Groups** > **Configure Online on Same Node Dependency > Change/Show Online on Same Node Dependency between Resource Groups** and press Enter.

   HACMP displays a list of resource groups configured with this location dependency.
3. Select the **Online on Same Node** dependency set of resource groups to show.
4. Add a resource group to the selected **Online on Same Node** dependency set of resource groups:

Resource Groups to be Online on the same node	HACMP displays the resource groups listed in the selected set.
New Resource Groups to be Online on the same node	Press F4 to display the list of available resource groups. Select the resource groups from the list to be in this set of resource groups to be acquired and brought ONLINE on the same node (according to the startup policy and the availability of the node required). On fallback and fallover, the resource groups are processed simultaneously and brought ONLINE on the same target node (using the fallover and fallback policy defined for these groups).

5. Press Enter.
6. Verify the cluster.

Changing an Online on Different Nodes Dependency

To change an **Online on Different Nodes** location dependency between resource groups:

1. Enter `smit hacmp`
2. In SMIT, select **HACMP Extended Resource Configuration > Configure Resource Group Run-Time Policies > Configure Dependencies between Resource Groups** > **Configure Online on Different Nodes Dependency > Change/Show Online on Different Nodes Dependency between Resource Groups** and press Enter.
3. Select the **Online on Different Nodes** dependency set of resource groups to show.
4. Make changes as required and then press Enter

High Priority Resource Group(s)	Select the resource groups to be in this set of resource groups to be acquired and brought ONLINE (according to the startup policy and the availability of the node required) before lower priority resource groups. On fallback and fallover, these resource groups are processed simultaneously and brought ONLINE on different target nodes before any other groups are processed. If different target nodes are unavailable for fallover or fallback, these groups (same priority level) can remain on the same node. Relative priority within this list is alphabetical by group names.

Intermediate Priority Resource Group(s)	Select the resource groups to be in this set of resource groups to be acquired and brought ONLINE (according to the startup policy and the availability of the node required) after the high priority groups and before low priority resource groups. are brought ONLINE. On fallback and fallover, these resource groups are processed simultaneously and brought ONLINE on different target nodes after the high priority groups and before low priority resource groups are processed. If different target nodes are unavailable for fallover or fallback, these groups (same priority level) can remain on the same node. Relative priority within this list is alphabetical by group names.
Low Priority Resource Group(s)	Select the resource groups to be in this set of resource groups to be acquired and brought ONLINE (according to the startup policy and the availability of the node required) after the higher priority resource groups are brought ONLINE. On fallback and fallover, these resource groups are processed and brought ONLINE on different target nodes after the higher priority groups are processed. . Higher priority groups moving to a node may cause these groups to be moved or taken OFFLINE. Relative priority within this list is alphabetical by group names.

5. Verify the cluster.

Changing an Online on Same Site Dependency

To change an **Online on Same Site** location dependency between resource groups:

1. Enter `smit hacmp`
2. In SMIT, select **HACMP Extended Resource Configuration > Configure Resource Group Run-Time Policies > Configure Dependencies between Resource Groups** > **Configure Online on Same Site Dependency > Change/Show Online on Same Site Dependency between Resource Groups** and press Enter.
3. Select the **Online on Same Site** dependency set of resource groups to show.
4. Add or remove resource groups from the list.
5. Verify the cluster.

Changing a Parent/Child Dependency between Resource Groups

To change a parent/child dependency between resource groups:

1. Enter `smit hacmp`

2. In SMIT, select **HACMP Extended Resource Configuration > Configure Resource Group Run-Time Policies > Configure Dependencies between Resource Groups** > **Configure Parent/Child Dependency >Change/Show Parent/Child Dependency between Resource Groups** and press Enter.

   A list of child-parent resource group pairs appears.

3. Select a pair from the list and press Enter. A screen appears where you can change the parent resource group or the child resource group.

4. Change the resource groups as required and press Enter. Note that you cannot change the **Dependency Type**.

Displaying a Parent/Child Dependency between Resource Groups

You can display parent/child dependencies between resource groups.

Note: You can use either ASCII SMIT or WebSMIT to display parent/child dependencies between resource groups. For more information on WebSMIT, see Chapter 2: Administering a Cluster Using WebSMIT.

To display a dependency between parent/child resource groups:

1. Enter `smit hacmp`

2. In SMIT, select **Extended Resource Configuration > Configure Resource Group Run-Time Policies > Configure Dependencies between Resource Groups** > **Configure Parent/Child Dependency > Display All Parent/Child Resource Group Dependencies** and press Enter.

   A selector screen appears.

3. Select **Display per Child** or **Display per Parent** to display all resource group dependencies for a child resource group, or for a parent resource group. Press Enter.

4. HACMP displays a list similar to one of the following:

```
Resource Group (RG_b) has the following parent resource
groups:
RG_a
RG_e
```

Or:

```
Resource Group (RG_a) has the following child resource
groups:
RG_b
RG_c
RG_d
Resource Group (RG_e) has the following child resource
groups:
RG_b
RG_c
RG_d
```

Displaying a Parent/Child Dependency between Resource Groups in WebSMIT

You can display parent/child dependencies between resource groups using WebSMIT.

Starting with HACMP 5.4, you can use WebSMIT to view graphical displays of sites, networks, nodes and resource group dependencies. For more information on WebSMIT, see Chapter 2: Administering a Cluster Using WebSMIT.

Removing a Dependency between Resource Groups

You can remove any of the four types of dependencies between resource groups.

Deleting a Parent/Child Dependency between Resource Groups

To delete a parent/child dependency between resource groups:

1. Enter `smit hacmp`
2. In SMIT, select **HACMP Extended Resource Configuration > Configure Resource Group Run-Time Policies > Configure Dependencies between Resource Groups** > **Configure Parent/Child Dependency > Delete a Dependency between Parent/Child Resource Groups** and press Enter.

   HACMP displays a list of child-parent resource group pairs.
3. Select a pair from the list to delete and press Enter. Deleting a dependency between resource groups *does not* delete the resource groups themselves.

   Note: If you have the following dependency chain configured: A > B > C, and remove the resource group B, HACMP sends a warning that the dependency links between A and B, and between B and C are also removed.

Deleting a Location Dependency between Resource Groups

To delete a location dependency between resource groups:

1. In SMIT, select the path for configuring the location dependency that you want to remove.

This example shows the path for Online on Same Node Dependency: **Extended Resource Configuration > Configure Resource Group Run-Time Policies > Configure Dependencies between Resource Groups** > **Configure Online on same node Dependency > Remove Online on Same Node Dependency between Resource Groups** and press Enter.

HACMP displays a list of resource groups with this location dependency.

2. Select the **Online on same node** dependency to remove and press Enter.

   Deleting a dependency between resource groups *does not* delete the resource groups themselves. The resource groups are now handled individually according to their site management, startup, fallover, and fallback policies.

Adding or Removing Individual Resources

You can add a resource to or remove a resource from a resource group in an active cluster without having to stop and restart cluster services to apply the change to the current configuration. You can add or remove resources from resource groups even if another node in the cluster is inactive. However, it is more convenient to have nodes active, so you can obtain a list of possible shared resources for each field by pressing the F4 key when you are in the SMIT **Change/Show Resources/Attributes for a Resource Group** panel.

Resource groups can contain many different types of cluster resources, including IP labels/addresses, filesystems, volume groups and application servers. You can change the mix of resources in a resource group and the settings of other cluster resource attributes by using the SMIT **Change/Show Resources/Attributes for a Resource Group** panel. See the following section.

Reconfiguring Resources in a Resource Group

To change the resources in a resource group:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Resource Configuration > Configure Resource Groups** > **Change/Show Resources/Attributes for a Resource Group** and press Enter. SMIT displays a picklist of configured resource groups.
3. Select the resource group you want to change and press Enter. SMIT displays a panel that lists all the types of resources that can be added to the type of selected resource group, with their current values.

   Note: If you specify filesystems to NFS mount in a non-concurrent resource group with the startup policy of either Online on Home Node Only, or Online on First Available Node, you must also configure the resource to use IP Address Takeover. If you do *not* do this, takeover results are unpredictable.You should also set the field value **Filesystems Mounted Before IP Configured** to **true** so that the takeover process proceeds correctly.

4. Enter the field values you want to change, and press Enter.
5. Return to previous SMIT panels to perform other configuration tasks or to synchronize the changes you just made.

If the Cluster Manager is running on the local node, synchronizing cluster resources triggers a dynamic reconfiguration event. For more information, see Chapter 7: Verifying and Synchronizing an HACMP Cluster.

Forcing a Varyon of a Volume Group

You can force a varyon of a volume group either by specifying an attribute in SMIT, or by entering a command at the command line. It is recommended that you use SMIT to force a varyon because HACMP does the following before attempting to activate a volume group on a node:

- Checks whether the LVM mirroring is used for the disks
- Verifies that at least one copy of every logical volume for this volume group can be found.

It is recommended to specify the **super strict** allocation policy for the logical volumes in volume groups for which forced varyon is specified.

As with regular volume group operations, you can determine the final status of the volume group using the messages logged by HACMP during the verification process and the information logged in the **hacmp.out** file. You can also use the **lsvg -o** command to verify whether a volume group is offline or online, and the **lsvg -l** command to check the volume group status and attributes.

If after checking partition maps, HACMP cannot find a complete copy of every logical volume for a volume group, an error message: "`Unable to vary on volume group <vg name> because logical volume <logical volume name> is incomplete`" displays in the **hacmp.out** file and the volume group remains offline.

For more information about the forced varyon functionality and quorum issues, see Chapter 5: Planning Shared LVM Components in the *Planning Guide.*

Forcing a Varyon of a Volume Group from SMIT

Note that you specify a forced varyon attribute for *all* volume groups that belong to a resource group. For instructions on setting a forced varyon attribute using SMIT, see Forcing a Varyon of Volume Groups in Chapter 5: Configuring HACMP Resource Groups (Extended).

With this attribute, if a normal **varyonvg** fails, a check is made to ensure that there is at least one complete copy of all data available in the volume group. If there is, it runs **varyonvg -f**; otherwise, the volume group remains offline. Specifying the forced varyon attribute for a volume group eliminates the need for a quorum buster disk or special scripts to force a varyon, although you can continue to use these methods.

Use the following procedure to ensure that you always have access to your data if there is a copy available, and that you receive notification if you lose either a copy of your data, or all copies.

To use HACMP forced varyon and error notification:

1. Disable quorum on your volume group. This will ensure that it does *not* vary off if you still have access to a copy of your data.
2. Use the SMIT forced varyon option to vary on your volume group if your data is available.
3. Set up error notification to inform you if a filesystem or logical volume becomes unavailable.

For information about creating scripts for cluster events, see Chapter 7: Planning for Cluster Events in the *Planning Guide*.

Forcing a Varyon of a Volume Group from the Command Line

Issue the **varyonvg -f** command for a specific volume group on a node in the cluster. If you use this method, HACMP does *not* verify that the disks are LVM mirrored, and does *not* check the logical partitions to verify that at least one complete copy of every logical volume can be found for this volume group. You should use this command with caution to avoid forcing a varyon of a volume group in a partitioned cluster. For more information, see the section Avoiding a Partitioned Cluster.

WARNING: Forcing a varyon with non-mirrored logical volumes *and* missing disk resources can cause unpredictable results (both conditions must be present to cause problems.) Forcing a varyon should only be performed with a complete understanding of the risks involved. For more information, see the following section. Also, refer to the AIX 5L documentation.

Avoiding a Partitioned Cluster

Use care when using forced varyon to activate a volume group. If the cluster becomes partitioned, each partition might force the volume group to vary on and continue to run. In this case, two different copies of the data are active at the same time. This situation is referred to as *data divergence*, and does *not* allow a clean recovery. If this happens in a concurrent volume group, the two sides of the cluster have made uncoordinated updates.

To prevent cluster partitioning, you should configure multiple heartbeat paths between the disks in the cluster.

Resource Group Migration

The Resource Group Management utility lets you perform maintenance on a node without losing access to the node's resources. You are *not* required to synchronize cluster resources or stop cluster services.

The Resource Group Management utility provides improved cluster management by allowing you to:

- Bring a resource group online or offline.
- Move a non-concurrent resource group to a new location. This location can be a node in the same site or a node in the other site.
  - *Non-concurrent resource groups* are resource groups that do *not* have the Online on All Available Nodes startup policy that is, they are *not* online simultaneously on multiple nodes in the cluster.
  - *Concurrent resource groups* are resource groups that have the Online on All Available Nodes startup policy that is, they start on *all* nodes in the cluster.

If you have requested HACMP to move, activate, or stop a particular resource group, then no additional operations on any additional groups will run until this operation is completed.

Specific considerations related to resource group migrations are:

- HACMP attempts to recover resource groups in the ERROR state upon **node_up** events. However, if you moved a group to Node A, the group remains on Node A (even in the error state).
- When node B joins the cluster, it does *not* acquire any resource groups that are currently in the ERROR state on node A. To recover such resource groups, manually bring them online or move them to other nodes.
- For a summary of **clRGmove** actions that are allowed in clusters with sites, see the section Migrating Resource Groups with Replicated Resources later in this chapter.

Note: When you request HACMP to perform resource group migration, it uses the **clRGmove** utility, which moves resource groups by calling an **rg_move** event. It is important to distinguish between an **rg_move** event that is triggered automatically by HACMP, and an **rg_move** event that occurs when you explicitly request HACMP to manage resource groups for you. To track and identify the causes of operations performed on the resource groups in the cluster, look for the command output in SMIT and for the information in the **hacmp.out** file.

This section covers:

- Requirements before Migrating a Resource Group
- Migrating Resource Groups with Dependencies
- Migrating Resource Groups Using SMIT
- Migrating Resource Groups from the Command Line
- Special Considerations when Stopping a Resource Group
- Checking Resource Group State
- Migrating Resource Groups with Replicated Resources.

Requirements before Migrating a Resource Group

Before attempting to explicitly move a resource group from one node to another, or to take a resource group online or offline, ensure that:

- HACMP 5.4 is installed on all nodes.
- The Cluster Manager is running on the node that releases the resource group and on the node that acquires it.
- The cluster is stable. If the cluster is *not* stable, the operation that you request with the resource group terminates and you receive an error message.

Migrating Resource Groups with Dependencies

HACMP prevents you from moving resource groups online or to another node under the following conditions:

- If you took the parent resource groups offline with the Resource Group Management utility, **clRGmove**, HACMP rejects manual attempts to bring the resource groups that depend on these resource groups online. The error message lists the parent resource groups that you must activate first to satisfy the resource group dependency.

- If you have a parent and a child resource group online, and would like to move the parent resource group to another node or take it offline, HACMP prevents you from doing so before a child resource group is taken offline. However, if both parent and child are in the same **Same Node** or **Same Site** location dependency set, you can move them both as you move the whole set.
- You can move **Same Node** dependency or **Same Site** dependency *sets* of resource groups. If you move one member of one of these sets, the whole set moves.
- The rules for location dependencies may *not* allow some moves.

See the section Moving Resource Groups with Dependencies in a Cluster with Sites.

Migrating Resource Groups Using SMIT

You can access the resource group migration functions using the SMIT interface as well as the command line. This section includes instructions on how to move, bring online, or take offline concurrent and non-concurrent resource groups using SMIT.

Note: You can access the resource group migration functions using the ASCII SMIT, WebSMIT, or the command line interface. This section includes instructions on how to move, bring online, or take offline concurrent and non-concurrent resource groups using ASCII SMIT.

For more information on the command line, see Migrating Resource Groups from the Command Line in this chapter.

For more information on WebSMIT, see Chapter 2: Administering a Cluster Using WebSMIT.

To manage resource groups in SMIT:

1. Enter `smit cl_admin`
2. In SMIT, select **HACMP Resource Group and Application Management** and press Enter.

   SMIT presents the following options for resource group migration.

Show Current State of Applications and Resource Groups	Displays the current states of applications and resource groups for each resource group. • For non-concurrent groups, HACMP shows only the node on which they are online and the applications state on this node. • For concurrent groups, HACMP shows ALL nodes on which they are online and the applications states on the nodes. • For groups that are offline on all nodes, only the application states are displayed; node names are *not* listed.

Bring a Resource Group Online	This option brings a resource group online by calling the **rg_move** event. Select this option to activate all resources in a specified resource group on the destination node that you specify, or on the node that is the current highest priority node for that resource group. For more information, see Bringing a Non-Concurrent Resource Group Online with SMIT or Bringing a Concurrent Resource Group Online with SMIT.
Bring a Resource Group Offline	This option brings a resource group offline on a node. Use this option to deactivate all resources in a specified resource group on a specific destination node. For more information, see Taking a Concurrent Resource Group Offline or Taking a Non-Concurrent Resource Group Offline with SMIT.
Move a Resource Group to another node/site	This option moves resource group(s) between nodes within the same site, or between sites. This option is applicable only to non-concurrent resource groups. For more information, see the section Moving Non-Concurrent Resource Groups to Another Node with SMIT.

3. Select a type of resource group migration you want to perform and press Enter.

Note: If you have replicated resources defined in your cluster, resource group selector panels and node selector panels in SMIT contain additional information to help you in the selection process.

- In resource group selector panels, the resource group state, owner node and site appear to the right of the resource group name.
- In node selector panels, if the node belongs to a site, the name of that site appears to the right of the node name.
- Only the nodes within the same site appear in the selector panel for the destination node.

Moving Non-Concurrent Resource Groups to Another Node with SMIT

You can move a non-concurrent resource group to a specified node in the cluster. You have two choices:

- Moving a non-concurrent resource group to a node at the same site with SMIT
- Moving a non-concurrent resource group to a node at another site with SMIT.

If you have defined location dependencies between resource groups (**Online On Same Node, Online At Same Site**) the picklist of resource groups lists these sets together on the same line. Moving any one of these resource groups causes the whole set to move.

If you have defined an **Online on Different Nodes** location dependency, HACMP may prevent you from moving a resource group or a set to a node if a higher priority resource group is already online on that node.

Moving a non-concurrent resource group to a node at the same site with SMIT

To move a non-concurrent resource group to another node at the same site:

1. Enter `smit cl_admin`
2. In SMIT, select **HACMP Resource Group and Application Management > Move a Resource Group to Another Node/Site** and press Enter.
3. Select **Move to another node within the same site** and press Enter.

   A picklist with resource groups appears. The list includes only those non-concurrent resource groups that are online on any node in the cluster.

   Note: Resource groups with Online On The Same Node dependencies are listed as sets together (to move the whole set).

4. Select the resource group from the picklist and press Enter. The **Select a Destination Node** panel with potential destination nodes appears.
5. SMIT displays the nodes on which the Cluster Manager is running, that are included in the resource group's nodelist, and that can acquire the resource group or set of groups:

Select a Destination Node	**Node Name**. Select one of the destination nodes. If HACMP finds that originally configured highest priority node is available, it places an asterisk (*) next to this node, as shown in the example below:

```
Select a Destination Node:

#* Denotes Originally Configured Highest Priority Node

* Node A
  Node C
```

   Press Enter.The next panel displays your selections.

6. Enter field values as follows:

Resource Groups to be Moved	Resource group(s) that will be moved.
Destination Node	Destination node that you selected in the previous panel. This field is non-editable.

7. Confirm your selections and press Enter. You do *not* need to synchronize the cluster. The **clRGmove** utility waits until the **rg_move** event completes. This may result in a successful migration of the resource group, or a failure. If the move fails, check **hacmp.out** for the reason.

Moving a non-concurrent resource group to a node at another site with SMIT

To move a non-concurrent resource group to a node at another site:

1. Enter `smit cl_admin`

2. In SMIT, select **HACMP Resource Group and Application Management > Move Resource Groups to Another Node/Site > Move Resource Groups to Another Site** and press Enter.
3. A picklist with resource groups appears. The list includes only those non-concurrent resource groups that are online on any node in the cluster. If you have defined dependencies between resource groups, these resource groups are listed on the same line in the same order you entered them when you defined the dependency.

   Resource groups with **same site** dependencies are listed first, followed by resource groups with **same node** dependencies, then individual resource groups.
4. Select the resource group (or set) from the picklist and press Enter. The **Select a Destination Site** panel appears. If HACMP finds that an originally configured Primary site for the group is now available to host the group, it indicates this in the picklist with an asterisk (*). Here is an example:

   ```
   Select a Destination Site:

   #* Denotes Originally Configured Primary Site
   * Site A
     Site B
   ```
5. Select a node at the destination site and press Enter.

   The **Move a Resource Group to Another Site** appears. Enter field values as follows:

Resource Groups to be Moved	Resource group(s) you selected.
Destination Site	Site you selected.

6. Confirm your selections and press Enter. You do *not* need to synchronize the cluster. The **clRGmove** utility waits until the **rg_move** event completes. This may result in a successful migration of the resource group, or a failure.

   If the event completes successfully, HACMP displays a message and the status and location of the resource group that was successfully moved. For an example of such output, see the section Using the clRGinfo Command in Chapter 10: Monitoring an HACMP Cluster.

   If the resource group migration fails, HACMP displays a message with the cause of failure. Be sure to take action in this case to stabilize the cluster, if needed. For more information, see the section No Automatic Recovery for Resource Groups That Fail to Migrate.

Bringing a Non-Concurrent Resource Group Online with SMIT

To bring a non-concurrent resource group online:

1. Enter `smit cl_admin`
2. In SMIT, select **HACMP Resource Group and Application Management > Bring a Resource Group Online**. The picklist appears. It lists resource groups that are offline or in the ERROR state on all nodes in the cluster.

3. Select the non-concurrent resource group from the picklist and press Enter.

   The **Select a Destination Node** picklist appears. It displays only the nodes that have cluster services running, participate in the resource group nodelist, and have enough available resources to host the resource group. The nodes in this list appear in the same order of priority as in the resource group nodelist.

4. Select a destination node using one of the following options and press Enter if HACMP finds that an originally configured highest priority node for the group is now available to host the group, it indicates this in the picklist with an asterisk (*). Here is an example:

   ```
   Select a Destination Node:

   #* Denotes Originally Configured Highest Priority Node
   * Node A
     Node B
   ```

   After selecting a destination node, the panel **Bring a Resource Group Online** appears.

5. Enter field values as follows:

Resource Group to Bring Online	Resource group to be activated.
Destination Node	Destination node you selected.

6. Confirm your selections and press Enter to start the execution of the **rg_move** event and bring the resource group online. You do *not* need to synchronize the cluster.

If the event completes successfully, HACMP displays a message and the status and location of the resource group that was successfully brought online on the specified node. For an example of such output, see the section Using the clRGinfo Command in Chapter 10: Monitoring an HACMP Cluster.

If you requested HACMP to activate the resource group on a particular node, and this node fails to bring the resource group online, the resource group is put into the ERROR state. In this case, HACMP does *not* attempt to activate the resource group on any other node in the cluster without your intervention. The error message in this case indicates that your intervention is required to activate the resource group on another node and stabilize the cluster.

Taking a Non-Concurrent Resource Group Offline with SMIT

To take a non-concurrent resource group offline:

1. Enter smit `cl_admin`
2. In SMIT, select **HACMP Resource Group and Application Management > Bring a Resource Group Offline**. The picklist appears. It lists only the resource groups that are online or in the ERROR state on all nodes in the cluster.
3. Select the non-concurrent resource group from the picklist and press Enter.

   After selecting a resource group, the **Select a Destination Node** picklist appears. It lists only the nodes on which cluster services are running and on which the resource group is currently online or in the ERROR state.

4. Select a destination node from the picklist. When you select it, it becomes a temporarily set highest priority node for this resource group.

After selecting a destination node, the panel **Bring a Resource Group Offline** appears.

5. Enter field values as follows:

Resource Group to Bring Offline	Resource group that is stopped or brought offline.
Destination Node	Node on which the resource group will be stopped.

6. Confirm your selections and press Enter to start the execution of the **rg_move** event and bring the resource group offline. You do *not* need to synchronize the cluster.

If the event completes successfully, HACMP displays a message and the status and location of the resource group that was successfully stopped on the specified node. For an example of such output, see the section Using the clRGinfo Command in Chapter 10: Monitoring an HACMP Cluster.

If you requested to bring a resource group offline on a particular node, and the resource group fails to release from the node on which it is online, an error message indicates that your intervention is required to stabilize the cluster.

Bringing a Concurrent Resource Group Online with SMIT

You can use SMIT to bring a concurrent resource group online either on one node or on ALL nodes in the cluster.

Note: *Concurrent resource groups* are those resource groups that have the Online on All Available Nodes startup policy; that is, they start on *all* nodes in the cluster.

Bringing a resource group online in SMIT or through the command line activates (starts) the specified resource group on one node, or on ALL nodes.

To bring a concurrent resource group online:

1. Enter `smit cl_admin`
2. In SMIT, select **HACMP Resource Group and Application Management > Bring a Resource Group Online**. The picklist appears. It lists only the resource groups that are offline or in the ERROR state on some or all nodes in the cluster.
3. Select the concurrent resource group from the picklist and press Enter.

   The **Select a Destination Node** picklist appears. HACMP displays only the nodes that have cluster services running, participate in the resource group's nodelist and have enough available resources to host the resource group. If HACMP finds that an originally configured highest priority node for the group is now available to host the group, it indicates this in the picklist with an asterisk (*).

4. Select a destination node and press Enter:

Select a Destination Node	**All_Nodes_in_Group**. If you select this option, the resource group will be activated on all nodes in the nodelist. In addition, selecting this option returns the resource group to its default behavior. **Node Name**. One of the destination nodes on which you can activate the resource group.

After selecting a destination node, the panel **Bring a Resource Group Online** appears. It shows your selections:

Resource Group to Bring Online	Specifies the resource group to be activated.
Destination Node	Specifies the destination node on which the resource group will be activated. If the **All_Nodes_in_Group** option is selected, the resource group gets started on *all* nodes in the nodelist.

5. Confirm your selections and press Enter to start the execution of the **rg_move** event and bring the resource group online. You do *not* need to synchronize the cluster.

If the event completes successfully, HACMP displays a message and the status and location of the resource group that was successfully brought online. For an example of such output, see the section Using the clRGinfo Command in Chapter 10: Monitoring an HACMP Cluster.

If you requested HACMP to activate the resource group on the node(s), and a particular node fails to bring the resource group online, the resource group is put into the ERROR state. In this case, HACMP does *not* attempt to activate the resource group on any other node in the cluster without your intervention. The error message in this case indicates that your intervention is required to activate the resource group on another node and stabilize the cluster.

Taking a Concurrent Resource Group Offline

You can use SMIT to take a concurrent resource group offline either on one node or on ALL nodes in the cluster.

When taking a resource group offline in SMIT or through the command line, you can select whether to make the offline state persist after rebooting the cluster, or *not*.

To take a concurrent resource group offline:

1. Enter `smit cl_admin`
2. In SMIT, select **HACMP Resource Group and Application Management > Bring a Resource Group Offline**. The picklist appears. It lists only the resource groups that are online or in the ERROR state on at least one node in the resource group nodelist.
3. Select the concurrent resource group from the picklist and press Enter.

   The **Select an Online Node** picklist appears. It displays a list of nodes that have cluster services running, and that appear in the same order of priority as in the resource group nodelist.

4. Select a destination node from the picklist and press Enter:

Select an Online Node	**All_Nodes_in_Group**. If you select this option, the resource group will be taken offline on ALL nodes in the nodelist. **Node Name**. One of the destination nodes on which you can take the resource group offline.

After selecting node(s), the panel **Bring a Concurrent Resource Group Offline** appears.

5. Enter field values as follows:

Resource Group to Bring Offline	Specifies the resource group to be brought offline.
Node on which to Bring Resource Group Offline	Specifies the destination node(s) on which the resource group will be brought offline. If the **All_Nodes_in_Group** option is selected, the resource group will be taken offline on ALL nodes. It will start on all nodes after you stop and restart the cluster services.

6. Press Enter to start the execution of the **rg_move** event to take the resource group offline. You do *not* need to synchronize the cluster.

If the event completes successfully, HACMP displays a message and the status and location of the resource group that was stopped on the node(s) in the cluster. For an example of such output, see the section Using the clRGinfo Command in Chapter 10: Monitoring an HACMP Cluster.

If you requested to bring a resource group offline on a particular node, and the resource group fails to release from the node on which it is online, an error message indicates that your intervention is required to stabilize the cluster.

No Automatic Recovery for Resource Groups That Fail to Migrate

If you request HACMP to move a resource group to a node and during this operation the destination node fails to acquire the group, the resource group is put into an ERROR state. If you try to move a resource group that has a dependency (parent/child or location) that prohibits the move, the resource group will be in the DEPENDENCY_ERROR state.

Similarly, if you request HACMP to activate the resource group on a particular node, and this node fails to bring the resource group online, the resource group is put into an ERROR state.

In either case, HACMP does *not* attempt to acquire or activate the resource group on any other node in the cluster. The error messages in these cases indicate that your intervention is required to move the resource group to another node.

If you request HACMP to migrate a resource group to another node, but the node that owns it fails to release it, or if you request to bring a resource group offline on a particular node, but the node fails to release it, an error message indicates that your intervention is required to stabilize the cluster.

Returning Previously Moved Resource Groups to Their Originally Configured Highest Priority Nodes

This section applies only to resource groups with:

- Fallback to Highest Priority Nodes fallback policies
- Resource groups that are configured to fall back, according to timers, for instance
- Fallback to Primary Site (if sites are defined) site fallback policies.

If you move such a resource group to a node other than its highest priority node (or to a site other than its Primary site) and the resource group is normally set to fall back to its highest priority node (or Primary site), then, after you moved it, it will fall back to the “new” node or site, *not* to the originally set highest priority node or Primary site.

Therefore, you may want to move this group back to the originally configured highest priority node, or Primary site, if they become available. You do so using the same SMIT panels as for moving a resource group and by selecting a node from the node list that has an asterisk (this indicates that this node is originally configured highest priority node). The same is true for sites.

These actions restore the highest priority node or Primary site for resource groups that you previously manually moved to other nodes. From this point on, these groups will continue to fall back to the highest priority node (or to a node at a Primary site).

Migrating Resource Groups from the Command Line

This section provides information about resource group migrations from the command line using the **clRGmove** command. For information on how to migrate a resource group using SMIT, see the section Requirements before Migrating a Resource Group. Before performing either method of resource group migration, you should read the preceding overview sections.

Note: You can access the resource group migration functions using the ASCII SMIT, WebSMIT, or the command line interface. This section includes instructions on how to move, bring online, or take offline concurrent and non-concurrent resource groups using the command line.

For more information on ASCII SMIT, see Migrating Resource Groups Using SMIT in this chapter.

For more information on WebSMIT, see Chapter 2: Administering a Cluster Using WebSMIT.

For full information on the **clRGmove** command and all of its associated flags, see the **clRGmove** man page in Appendix A: Script Utilities in the *Troubleshooting Guide*.

The **clRGmove** utility lets you manually control the location and the state of resource groups by calling the **rg_move** event. With this command you can bring a specified resource group offline or online, or move a resource group to a different node. This utility provides the command line interface to the Resource Group Migration functionality, which can be accessed through SMIT. You can also use this command from the command line, or include it in the pre- and post-event scripts.

In this section, the phrase “non-concurrent resource group” refers to resource groups with a startup policy that is *not* **Online On All Available Nodes**. The phrase concurrent resource group is used to refer to a resource group with the startup policy of **Online On All Available Nodes**.

For a non-concurrent resource group, you can:

- Take the resource group offline from an online node
- Bring the resource group online to a specific node
- Move the resource group from its current hosting node to a new location.

For a concurrent resource group, you can:

- Take the resource group offline from all nodes in the group's nodelist
- Take the resource group offline from one node in the group's nodelist
- Bring the resource group online on all nodes in the group's nodelist
- Bring the resource group online on one node in the group's nodelist.

Example: Using clRGmove to Swap Resource Groups

In the three-node cluster indicated here, each node—Node1, Node2, and Node3—has a service label/IP address and a standby label/IP address. There are three non-concurrent resource groups with the following policies:

- Startup: **Online on Home Node Only**
- Fallover: **Fallover to Next Priority Node in the List**
- Fallback: **Fallback to Higher Priority Node in the List**.

These resource groups have node priority lists as follows:

RG1 - Node1, Node3 **CrucialRG - Node2, Node3** **RG3 - Node3, Node1**

Each node is up and possesses a resource group as follows:

Node1 - UP (RG1) **Node2 - UP (CrucialRG)** **Node3 - UP (RG3)**

Node2’s resources—contained in **CrucialRG**—are of particular importance to your operation. A situation occurs in which two cluster nodes fail. Node1 fails first; its resources fall over to Node3, since Node3 is in RG1’s priority list. Then Node2 fails. In this case, Node2’s crucial resources remain down; they have nowhere to go, since Node3’s only standby label/IP address has already been taken. The cluster now looks like this:

Node1 - DOWN **Node2 - DOWN** **Node3 - UP (RG3, RG1)**

The crucial resource group is unavailable. HACMP is able to take care of only one failure, because there are no more standby label/IP addresses, so it handles the first failure, Node1, but *not* the second. However, if you need **CrucialRG**’s resources more than you need `RG1's`, you can use the Resource Group Management utility to “swap” the resource groups so you can access **CrucialRG** instead of `RG1`.

You do this by issuing the following commands:

```
clRGmove -g RG1 -n node3 -d
```

to bring `RG1` offline on `Node3`, and

```
clRGmove -g CrucialRG -n node3 -u
```

to bring **CrucialRG** online on `Node3`.

For more information, see the reference page for the **clRGmove** in Appendix A: Script Utilities in the *Troubleshooting Guide*.

After these resource group migration commands are completed, access to `CrucialRG` is restored, and the cluster looks like this:

Node1 - DOWN	**Node2 - DOWN**	**Node3 - UP (RG3, CrucialRG)**

Note: Only one resource group may be moved at a time with **clRGmove**.

Special Considerations when Stopping a Resource Group

After taking a resource group offline, you should *not* assume that a joining or rejoining node will bring that resource group online. The following are instances when a resource group must be brought back online using the Resource Group and Application Management utility.

- If you use **clRGmove -d** to bring down a resource group with Online on Home Node startup policy, Fallover to Next Priority Node in the List fallover policy and Fallback to Higher Priority Node in the List fallback policy, and which resides on the highest priority node, it will remain in an inactive state. You must manually bring the resource group online through resource group management.
- If you specify the **fallover** option of application monitoring for a resource group using the **Customize Resource Recovery** SMIT panel, which may cause resource groups to migrate from their original owner node, the possibility exists that while the highest priority node is up, the resource group remains down. Unless you bring the resource group up manually, it will remain in an inactive state.
- If your resource group was placed in an UNMANAGED state, due to stopping cluster services without stopping the applications, you may need to bring this resource group online manually.

See Chapter 3: Investigating System Components and Solving Common Problems in the *Troubleshooting Guide* for more information.

Checking Resource Group State

As with regular cluster events, you can debug the status of resource groups using the messages logged by HACMP in the **hacmp.out** file.

In addition, you can use **clRGinfo** to view the resource group location and status. See the section Using the clRGinfo Command in Chapter 10: Monitoring an HACMP Cluster for an example of the command output. Use **clRGinfo -p** to view the node that is temporarily the highest priority node.

Migrating Resource Groups with Replicated Resources

Starting with HACMP 5.3, the Resource Group Management utility provides additional support for moving resource groups that contain replicated resources. If you have sites defined in your cluster, you can take a resource group offline, online, or move it to another node *in either site* in the cluster. You use the same SMIT panels to perform these operations as you use in a cluster without sites.

If sites are defined in the cluster, the resource group can be in one of the following states:

On the Primary Site	On the Secondary Site
ONLINE	ONLINE SECONDARY
OFFLINE	OFFLINE
ERROR	ERROR SECONDARY
UNMANAGED	UNMANAGED SECONDARY

Resource group states depend upon the **Inter-Site Management Policy** defined for a resource group.

Whatever actions a resource group in the ONLINE or in the ONLINE SECONDARY state is allowed to perform (be acquired on a node, or mount filesystems) depends on the type of replicated resources you configure in the cluster.

You can use **clRGmove** to move a resource group online, offline, or to another node within a site, or to a node on the other site. See Migrating Resource Groups from the Command Line for more information.

If you have configured resource group dependencies, these may limit possible user actions. See Migrating Resource Groups with Dependencies for more information.

If you have configured replicated resources in your cluster, you can perform the following resource group migrations:

On the Primary Site:

ONLINE > ONLINE Move a resource group to another node within the same site.	The resource group will return to its default behavior after a cluster reboot.
OFFLINE > ONLINE **ERROR > ONLINE** Bring online on a node in primary site.	If the resource group is in ONLINE_SECONDARY at this site, HACMP first moves the secondary instance to the other site.
ONLINE > OFFLINE **ERROR > OFFLINE** Bring offline on a node.	Note that in this case, the resource group may remain in the ONLINE SECONDARY state on a node in the secondary site.

On the Secondary Site:

ONLINE SECONDARY > ONLINE SECONDARY Move a resource group to another node within the same site.	Resource group will return to its default behavior after a cluster reboot.
OFFLINE SECONDARY or **ERROR SECONDARY > ONLINE SECONDARY** Bring online on a node.	HACMP lets you select only those resource groups that are *not* in the ONLINE state on the site.
ONLINE SECONDARY or ERROR SECONDARY > OFFLINE Bring offline on a node.	The resource group may remain online on a node on the primary site. =

Migrating between Sites in a Non-concurrent Node Policy /Non-concurrent Site Policy Configuration

Primary Instance > Other Site Move a resource group to a node on the other site.	The secondary instance automatically moves to the opposite site to the highest priority node. The resource group will return to its default behavior after a cluster reboot.
OFFLINE or ERROR >ONLINE Bring primary instance online on a node on either site.	If the secondary instance is online on that site, HACMP releases the secondary instance before acquiring the primary instance. The secondary instance moves to the opposite site.
OFFLINE SECONDARY or **ERROR SECONDARY > ONLINE SECONDARY** Bring secondary instance online on either site.	If the primary instance is online at a site, the secondary instance cannot be brought online on the same site.
ONLINE SECONDARY > OFFLINE **ERROR SECONDARY> OFFLINE** Bring offline on a node.	The resource group may remain online on a node on the primary site.

Migrating between Sites in a Concurrent Node Policy/Non-concurrent Site policy Configuration

ONLINE > OFFLINE **OFFLINE > ONLINE** **or** **ONLINE_SECONDARY > OFFLINE_SECONDARY** **OFFLINE_SECONDARY > ONLINE_SECONDARY** Bring a resource group online or offline on particular node on a site.	The resource group is in ONLINE state on all nodes at the primary site, and is in ONLINE_SECONDARY state on all nodes at the other site. You can take particular nodes on or offline. This does *not* affect the instances on the other site.
Swap Sites Move primary instances to the other site; secondary instances move to opposite site.	The primary instance is always on the opposite site from the secondary instance (in this case, on all nodes).

For more information on configuring replicated resources in HACMP, see the *Planning Guide.*

Moving Resource Groups with Dependencies in a Cluster with Sites

The following table shows the various actions you can take to migrate resource groups and the interactions or limitations caused by resource group dependencies in the cluster.:

User Migration Action	Limitations or Comments
Move an **Online On Same Nodes** dependency set within a site to a specific node.	If a resource group that belongs to an **Online on Different Nodes** dependency set is already online on the node, SMIT prevents this action.
Move a resource group within a site to a specific node.	If a resource group that belongs to an **Online on Different Nodes** dependency set is already online on the node, SMIT prevents this action.
Move an **Online On Same Nodes** dependency set across sites to a specific node.	If a resource group that belongs to an **Online on Different Nodes** dependency set is already online on the node, SMIT prevents this action.
Move an **Online On Same Nodes** dependency set across sites to a specific node.	If a resource group that belongs to an **Online on Different Nodes** dependency set is already online on the node, SMIT prevents this action.
Move an **Online On Same Nodes** dependency set across sites, specifying a *site* as the destination.	

User Migration Action	Limitations or Comments
Move an **Online On Same Nodes** dependency set within a site to a specific node.	If a resource group that belongs to an **Online on Different Nodes** dependency set is already online on the node, SMIT prevents this action.
Move an **Online on Same Site** dependency set across sites.	
Move a resource group across sites to a specific node.	If a resource group that belongs to an **Online on Different Nodes** dependency set is already online on the node, SMIT prevents this action.
Move a resource group across sites, specifying a *site* as destination	
Move all instances of a set of concurrent/non-concurrent resource groups to the opposite site.	In essence, the primary and the secondary instances will swap sites.
Move all instances of a concurrent/non-concurrent resource group to the opposite site.	In essence, the primary and the secondary instances will swap sites.
Move a secondary instance of a resource group within a site.	
Move a secondary instance of a resource group across sites, if the primary instance is *not* online on the target site.	

Bringing Resource Groups in a Cluster with Sites Offline

You can take primary or secondary instances of non-concurrent or concurrent resource groups offline. However, if you try to bring a *last* primary instance of a concurrent or non-concurrent resource group offline on a node, HACMP prevents this to avoid an unnecessary inter-site fallover:

Bringing Resource Groups in a Cluster with Sites Online

The following table shows the various actions you can take to bring resource groups online and the limitations to this action.

User Action	Limitations
Bring the primary instance of a non-concurrent resource group online on a node.	If the secondary instance is on the site of the target node, the secondary instance has to move to the opposite site automatically. If an **Online on Different Nodes** dependency resource group is already online on the node, SMIT prevents this action.
Bring the secondary instance of a non-concurrent resource group online on a node.	The primary instance cannot be online on the site of the target node, in order to prevent unintentional movement of the primary instance.
Bring the primary instance of a concurrent resource group online on a node.	
Bring the primary instance of a concurrent/non-concurrent resource group online on a node.	If the secondary instance(s) is on the site of the target node, the secondary instance(s) has to move to the opposite site automatically.
Bring the secondary instance of a concurrent/non-concurrent resource group online on a node.	The primary instance(s) cannot be online on the site of the target node, in order to prevent unintentional movement of the primary instance(s).

Migrating Replicated Resource Groups Using SMIT

Using the SMIT interface is recommended for these migrations.

Moving a Concurrent Replicated Resource Group to Another Site

You can move a concurrent resource group to the other site in the cluster through the **HACMP Resource Group Management > Move a Resource Group** SMIT path.In effect, you swap the primary and secondary instances of the resource group between the sites.

To move a concurrent resource group to another site:

1. Enter `smit cl_admin`

2. In SMIT, select **HACMP Resource Group and Application Management > Move a Resource Group**. A picklist with resource groups appears. The list includes only those resource groups that are online on any node in the cluster and have defined a site policy other than Ignore or Online on Both Sites.
3. Select the resource group from the picklist and press Enter. The **Select a Destination Site** panel appears showing potential destination sites.
4. SMIT displays the sites on which the Cluster Manager is running, which are included in the resource group's nodelist, and which can acquire the resource group.

Select a Destination Site	**Site Name**. If you select one of the destination sites, it becomes a hosting node for this resource group. **Swap Sites** (for concurrent replicated resource group). This swaps primary and secondary instances between sites.

After you have selected the destination node for the resource group, press Enter.

The next panel displays your selections.

5. Enter field values as follows:

Resource Group to be Moved	Specifies the resource group that will be moved.
Destination Site for Primary Instances	The site name is listed. (Opposite of current site.)
Destination Site for Secondary Instances	The site name is listed. (Opposite of current site.)

6. Confirm your selections and press Enter. You do *not* need to synchronize the cluster.

If the event completes successfully, HACMP displays a message and the status and location of the resource group that was successfully moved. For an example of such output, see the section Using the clRGinfo Command in Chapter 10: Monitoring an HACMP Cluster.

If the resource group migration fails, HACMP displays a message with the cause of failure. Be sure to take action in this case to stabilize the cluster, if needed. For more information, see the section No Automatic Recovery for Resource Groups That Fail to Migrate.

Customizing Inter-Site Resource Group Recovery

Starting with HACMP 5.3, inter-site resource group recovery is enabled by default for a *new* installation.

Selective fallover of resource groups between sites is disabled by default when you upgrade to HACMP 5.3 or 5.4 from a previous release. This is the pre-5.3 release behavior for non-Ignore site management policy. A particular instance of a resource group can fall over within one site, but cannot move between sites. If no nodes are available on the site where the affected instance resides, that instance goes into ERROR or ERROR_SECONDARY state. It does *not* stay on the node where it failed. This behavior applies to both primary and secondary instances.

Note that even if the Cluster Manager is *not* enabled to initiate a selective fallover across sites, it will still move the resource group if a **node_down** or **node_up** event occurs. You can manually move a resource group between sites.

Enabling or Disabling Selective Fallover between Sites

You can change the resource group recovery policy to allow or disallow the Cluster Manager to move a resource group to another site in cases where it can use selective fallover to avoid having the resource group go into ERROR state.

Inter-Site Recovery of Replicated Resource Groups

If you enable selective fallover across sites, HACMP tries to recover both the primary and the secondary instance of a resource group in these situations:

- If an acquisition failure occurs while the secondary instance of a resource group is being acquired, the Cluster Manager tries to recover the resource group's secondary instance, as it does for the primary instance. If no nodes are available for the acquisition, the resource group's secondary instance goes into global ERROR_SECONDARY state.
- If quorum loss is triggered, and the resource group has its secondary instance online on the affected node, HACMP tries to recover the secondary instance on another available node.
- If a **local_network_down** occurs on an **XD_data** or **Geo_primary** network, HACMP moves replicated resource groups that are ONLINE on the particular node that have GLVM or HAGEO resources to another available node on that site. This functionality of the primary instance is mirrored to the secondary instance so that secondary instances may be recovered via selective fallover.

Using SMIT to Enable or Disable Inter-Site Selective Fallover

To enable or disable the Resource Group Recovery with Selective Fallover behavior:

1. In SMIT, select **Extended Configuration > Extended Resource Configuration > Customize Resource Group and Resource Recovery > Customize Inter-site Resource Group Recovery** and press Enter.

   A selector screen lists the resource groups that contain nodes from more than one site (including those with a site management policy of **Ignore**. These are *not* affected by this function even if you select one of them.)

2. Select the resource groups for recovery customization.

   The next screen lists the selected resource groups and includes the field to enable or disable inter-site selective fallover.

3. To enable inter-site selective fallover (initiated by the Cluster Manager), select **true.** The default is **false** for a cluster migrated from a previous release, and **true** for a new HACMP 5.3 or 5.4 cluster.

Chapter 16: Managing User and Groups

This chapter describes how to use the SMIT System Management (C-SPOC) utility to manage user accounts and groups on all nodes in a cluster by making configuration changes on a single node.

The chapter include the following sections:

- Overview
- Managing User Accounts across a Cluster
- Managing Password Changes for Users
- Changing the Password for Your Own User Account
- Managing Group Accounts.

Overview

HACMP lets you manage AIX 5L user and group accounts across an HACMP cluster. Groups provide an additional level of security and enable system administrators to manipulate a group of users as a single entity. In addition, HACMP provides a utility that lets you authorize specified users to change their own password across nodes in an HACMP cluster.

Requirements for Managing User Accounts in an HACMP Cluster

AIX 5L files that store user account information should be consistent across cluster nodes. These files are:

- The system **/etc/passwd** file
- Other system files in the **/etc/security** directory.

This way if a cluster node fails, users can log on to the surviving nodes without experiencing problems caused by mismatched user or group IDs.

As the system administrator of an HACMP cluster, you can use the C-SPOC utility to manage user and group accounts from any node in a cluster. C-SPOC propagates new and updated information to all of the other nodes in the cluster.

Note: Managing user accounts through C-SPOC requires that the Cluster Communications daemon is running and that all cluster nodes are active.

WARNING: If you manage user accounts with a utility such as Network Information Service (NIS) or Distributed Computing Environment (DCE) Manager, do *not* use HACMP user management. Using HACMP user management in this environment might cause serious system inconsistencies in the database.

User Account Configuration

Make sure user accounts are the same on all nodes in the cluster. *Run verification after you make changes to user accounts*.

If a node in the cluster has fewer password restrictions than the other nodes, a user could make changes from the node with fewer restrictions and degrade cluster security.

Status of C-SPOC Actions

If an action initiated by the C-SPOC utility fails, check the C-SPOC log file, **/tmp/cspoc.log**, to obtain the status of the command on each cluster node.

Note: The default location of this log file is **/tmp/cspoc.log**. If you redirected this log, check the appropriate location.

Managing User Accounts across a Cluster

You can manage user accounts from any node in a cluster by:

- Listing Users On All Cluster Nodes
- Adding User Accounts on all Cluster Nodes
- Changing Attributes of User Accounts in a Cluster
- Removing User Accounts from a Cluster

Starting with HACMP version 5.2, you can authorize users to change their own password and have C-SPOC propagate that password across cluster nodes. For information about this feature, see the section Managing Password Changes for Users.

Listing Users On All Cluster Nodes

To obtain information about all user accounts on cluster nodes, or on the nodes in a specified resource group, you can use the following procedure or run the **cl_lsuser** command. For information about the **cl_lsuser** command, see its man page.

To list all user accounts on *all* cluster nodes using the C-SPOC utility:

1. Enter the fastpath `smit cl_admin`
2. In SMIT, select **HACMP Security and Users Management > Users in an HACMP Cluster > List Users in the Cluster.**
3. In the **List Users in the Cluster** panel, leave the selection for a resource group blank to display information about all users, and press Enter.

   When you press Enter, SMIT executes the **cl_lsuser** command and displays a listing of user accounts similar to the following:

```
                        COMMAND STATUS

Command: OK            stdout: yes           stderr: no

Before command completion, additional instructions may appear below.
```

```
[TOP]
sigmund  root   0       /
sigmund  daemon 1       /etc
sigmund  bin    2       /bin
sigmund  sys    3       /usr/sys
sigmund  adm    4       /var/adm
sigmund  uucp   5       /usr/lib/uucp
sigmund  guest  100     /home/guest
sigmund  nobody -2      /
sigmund  lpd    9       /
sigmund  nuucp  6       /var/spool/uucppublic
orion    root   0       /
orion    daemon 1       /etc
orion    bin    2       /bin
[MORE...18]
```

Adding User Accounts on all Cluster Nodes

In AIX 5L, you can add user accounts by using either:

- The **mkuser** command
- **smit mkuser**

The user account information is stored in the **/etc/passwd** file and the files in the **/etc/security** directory. For more information about the **mkuser** command, see its man page.

For information about managing user accounts in AIX 5L, see your AIX 5L documentation. You can locate the AIX 5L documentation from the following URL:

http://publib16.boulder.ibm.com/pseries/en_US/infocenter/base/aix.htm

To add a user to all nodes in a cluster using the C-SPOC utility, perform the following procedure on any cluster node:

1. Enter `smit cl_admin`
2. In SMIT, select **HACMP Security and Users Management > Users in an HACMP Cluster > Add a User to the Cluster** and press Enter.

3. Enter data in the entry fields to set up the account.

   AIX 5L provides help panels that describe each attribute. The **User Name** field is the only required field.

   Note: You should specify a value in the **User ID** field so that the account's user ID is the same on all cluster nodes. If you do *not* specify this value, AIX 5L could assign a different user IDs on each nodes. A mismatch of user IDs for an account could prevent a user from logging on to another cluster node in the event of a fallover.

4. After entering user data, press Enter. The user account specified is created on all cluster nodes.

The C-SPOC utility creates the AIX 5L user account and home directory for the new account on each remote cluster node that you specify.

If a user with the same name already exists on one of the cluster nodes, the operation fails, returning this message:

`user-name already exists on node nodename`

You can specify that the command continue processing even if the user name already exists on one of the cluster nodes by specifying the **force** option.

Changing Attributes of User Accounts in a Cluster

In AIX 5L, you can change any of the attributes associated with an existing user account by using either:

- The **chuser** command
- The AIX 5L SMIT **Change User Attributes** panel.

The **chuser** command modifies the user information stored in the **/etc/passwd** file and the files in the **/etc/security** directory. For more information about the **chuser** command, see its man page.

You can also change attributes associated with an existing user account from C-SPOC, as described in the following procedure. This procedure executes the AIX 5L **chuser** command on each cluster node. All cluster nodes must be active, the Cluster Communications daemon running, and a user with the specified name must exist on all the nodes for the change operation to proceed.

To change the characteristics of a user account on all cluster nodes using the C-SPOC utility:

1. Enter the fastpath `smit cl_chuser`

   SMIT displays the **Change/Show Characteristics of a User in the Cluster** panel:

```
   Change / Show Characteristics of a User in the Cluster

   Type or select a value for the entry field.
   Press Enter AFTER making all desired changes.

   [Entry Fields]
   Select nodes by Resource Group                  []
+
           *** No selection means all nodes! ***
```

2. Specify the name of the user account you want to change and press Enter. Press F4 to obtain a listing of users from which to select. SMIT displays a complete listing of the user account attributes with their current values filled in.
3. Enter the new values for attributes you want to change and press Enter. AIX 5L provides help panels that explain each attribute. SMIT executes the C-SPOC command to change the attributes of the user account on all cluster nodes.

Removing User Accounts from a Cluster

In AIX 5L, you remove a user account by using either:

- The **rmuser** command
- The fastpath **smit cl_rmuser**

For information about the **rmuser** command, see its man page.

You can also remove a user account from cluster nodes from C-SPOC, as described in the following procedure. This procedure executes the AIX 5L **rmuser** command on all cluster nodes.

Note: The system removes the user account but does *not* remove the home directory or any files owned by the user. These files are accessible only to users with root permissions or by the group in which the user was a member.

To remove a user account from all cluster nodes using the C-SPOC utility:

1. Enter the fastpath `smit cl_rmuser`

   SMIT displays the **Remove a User** panel.

2. Enter field data as follows:

User Name	Enter a user name. The user name can be up to 8 characters in length.
Remove Authentication information?	Specify **Yes** to delete the password and other authentication information from system security files.

3. Press Enter.

   SMIT removes the specified user account from cluster nodes.

Managing Password Changes for Users

You can manage user passwords from any node in a cluster by:

- Changing Passwords for User Accounts
- Allowing Users to Change Their Own Passwords

  Starting with HACMP version 5.2, you can let specified users change their password on multiple nodes in the cluster by changing their password on one node.

  Note: Changing user passwords is *not* supported on SP nodes that use PSSP user management.

An HACMP user, that is a user who has an AIX 5L user account on each node in a cluster, can use the C-SPOC utility to change their own password across nodes in the cluster. For information about how a user changes their own password, see the section Changing the Password for Your Own User Account.

WARNING: If you manage user accounts with a utility such as Network Information Service (NIS) or Distributed Computing Environment (DCE) Manager, do *not* use HACMP user management. Using HACMP user management in this environment might cause serious system inconsistencies in the database.

Prerequisites for Allowing Users to Change Passwords

Before you authorize users to change their password or change a user's passwords, ensure that:

- The cluster topology is configured properly.
- The user's account exists on every cluster node in a specified resource group, and if no resource group is specified, in the entire cluster.
- The user's account exists on the local node. (The password changes on the local node, even if that node is *not* in the selected resource group.)
- All cluster nodes are powered up and accessible.

Note: These conditions should also be met before a user changes their own password. As a user may *not* have this information, the utility displays messages to a user should their attempt to change their password fail.

Allowing Users to Change Their Own Passwords

In HACMP version 5.2 and up, system administrators can enable the new Cluster Password (**clpasswd)** utility. This utility, when enabled, links to the AIX 5L system password utility to:

- Let system administrators authorize specified users to change their password across cluster nodes
- Let authorized users change their own password across a resource group or cluster (as configured), rather than having to change their password on each node in the cluster.

  This means that the user's AIX 5L system password is the same on the set of nodes specified.

Note: The security of the password propagated to other nodes is only as secure as the network used to distribute the password.

Depending on the configuration of the Cluster Password utility, it lets users change their password from:

- C-SPOC as described in the section Changing the Password for Your Own User Account
- The **clpasswd** command.

Both of these call the AIX 5L **passwd** command. The **clpasswd** command uses the same arguments as the **passwd** command. For more information about the **clpasswd** command, see its man page.

The following table shows where a user's password is changed based on the user's authorization, the password utility that is active, and the command executed:

	When the system password utility is linked to clpasswd and the AIX 5L passwd command is run	**When the system password utility is active (not linked to clpasswd)**	
		The AIX 5L passwd command is run	**The HACMP clpasswd command is run**
The user authorized to change password across cluster	The password is changed on all cluster nodes.	The password is changed only on the local node.	The password is changed on all cluster nodes.
The *user is not* authorized to change password across cluster	The password is changed only on the local node.	The password is changed only on the local node.	The password is *not* changed.

Configuring the Cluster Password Utility

To enable the Cluster Password utility:

1. Enter `smit cl_admin`
2. In SMIT, select **HACMP Security and User Management > Passwords in an HACMP Cluster > Modify System Password Utility**.

   The **Modify System Password Utility** panel appears.

3. Enter field values as follows:

/bin/passwd utility is	Select **Link to Cluster Password Utility** to link the Cluster Password Utility to the AIX 5L password utility. This enables the Cluster Password utility. Select **Original AIX System Command** to remove the link from the Cluster Password utility to the AIX 5L password utility. This disables the Cluster Password utility.
Select Nodes by Resource Group	Select one or more resource groups to enable the Cluster Password utility on the nodes in the specified group(s). Leave the field blank, to enable the Cluster Password utility on all cluster nodes.

When the Cluster Password utility is linked to the AIX 5L password utility, HACMP creates a **/usr/es/sbin/cluster/etc/clpasswd/passwd.orig** file to store the AIX 5L **passwd** utility. If you disable the Cluster Password utility, HACMP removes the link between the two files, and the **passwd.orig** file is moved to **/bin/passwd**.

Configuring Authorization

After the Cluster Password utility is linked to the AIX 5L system password utility (**passwd**), you can specify and update which users have permission to change their passwords across a cluster. For information about linking the Cluster Password utility, see the section Configuring the Cluster Password Utility.

To specify which users can change their own password:

1. Enter `smit cl_admin`
2. In SMIT, select **HACMP Security and User Management > Passwords in an HACMP Cluster > Manage List of Users Allowed to Change Password** and press Enter.
3. In the **Manage List of Users Allowed to Change Password** panel, view a list of users and select which users you want to allow to change their password across cluster nodes.

   or

   Select **ALL** to allow all cluster users to change their password across the cluster.

   You can also view the list of users allowed to change their password across a cluster, and select and remove a user from the list.

The **/etc/clpasswd/cl_passwd_users** file stores the list of users allowed to change their password across a cluster.

Changing Passwords for User Accounts

As administrator, you can use C-SPOC to change users' passwords or to specify that particular users change their password on next login. You can direct that this change take place on all cluster nodes or on nodes that are part of specified resource groups.

If you use C-SPOC to change a user password for all nodes that belong to a resource group, make sure you perform this operation on a node that is included in the resource group. If you run this C-SPOC command from a node that is *not* part of the resource group, the password changes on that node also.

Note: All nodes must have HACMP version 5.2 or higher installed. For other prerequisites, see the section Prerequisites for Allowing Users to Change Passwords.

To use SMIT to change a user's password on a list of nodes in the cluster:

1. Enter the fastpath `smit cl_chpasswd`
2. In the **Change a User's Password in the Cluster** panel, select the resource group that contains the nodes on which the user has an account and press Enter.

   If you leave the field blank, all nodes in the cluster are selected.
3. Enter field values as follows:

User Name	Select the name of the user whose password you want to change.
User must change Password on first login?	Set to **true** to require the user to change the password on each node on the next login. Set to **false** if you do *not* want to require the user to change the password on the next login. The default is **true**.

4. Press Enter to change the password.

   The panels that appear are similar to the AIX 5L **Change a User's Password** panels. You enter the user name and the current password and then change the password.

Changing the Password for Your Own User Account

As an individual user, you can change your password on all cluster nodes, or on nodes within a specified resource group, if:

- The Cluster Password utility is enabled on each cluster node.
- The administrator (who has root privileges) has given you permission to change your password on nodes across a cluster.

Note: The password you are changing is your AIX 5L password on the specified nodes.

If you are unsure whether or *not* you are authorized to change your password, or if you try to change your password and receive an error message, contact your system administrator.

For information about how the configuration for the Cluster Password utility affects where your password is changed, see the section Allowing Users to Change Their Own Passwords.

To change your password on cluster nodes:

1. Enter `smit hacmp`
2. In SMIT, select **System Management (C-SPOC) > HACMP Security and User Management > Password in an HACMP Cluster > Change Current User's Password** and press Enter.

   The **Change Current User's Password** panel appears.
3. Enter field values as follows:

Select nodes by Resource Group	Select the resource group(s) that contains the nodes where you want to change your password. Leave the field blank to select all nodes in the cluster.
User Name	Verify that this field displays your user name. If it displays another name, contact your system administrator.

4. Press Enter.
5. Change your password on the panel that appears.

If C-SPOC can distribute your new password to all cluster nodes or the nodes in a specified resource group, it changes your password across the nodes. Messages advise you of the progress of the password change and display the nodes on which the change takes place.

If C-SPOC cannot communicate with all cluster nodes, it does *not* change your password, and it displays a message to that affect.

Note: If your password is changed on some, but *not* all, of the cluster nodes, a message appears that directs you to contact your system administrator. Be sure to talk with your system administrator because your password might be inconsistent among nodes in the specified resource groups or cluster.

You can also use the **clpasswd** command to change your cluster password. If you have *not* been authorized to change your password on cluster nodes, the **clpasswd** command does *not* let you change your password on any node, including the one you are currently logged in to.

Managing Group Accounts

All users must belong to a group. Groups add a level of security.

WARNING: If you manage user accounts with a utility such as Network Information Service (NIS) or Distributed Computing Environment (DCE) Manager, do *not* use HACMP user management. Using HACMP user management in this environment might cause serious system inconsistencies in the database.

You can manage group accounts from any node in a cluster by:

- Listing Groups on All Cluster Nodes
- Adding Groups on Cluster Nodes
- Changing Characteristics of Groups in a Cluster
- Removing Groups from the Cluster

Listing Groups on All Cluster Nodes

Each group has associated attributes that include the names of the users in the group, the user name of the administrator of the group, and the group ID. In AIX 5L, you obtain information about all the groups defined on an AIX 5L system by running the **lsgroup** command. For information about the **lsgroup** command, see its man page.

You can obtain information about the groups defined on all cluster nodes from C-SPOC, as described in the following procedure, or by running the C-SPOC **cl_lsgroup** command specifying the **ALL** argument. (For more information about the **cl_lsgroup** command, see its man page.) Both C-SPOC and the **cl_lsgroup** command execute the **lsgroup** command on each cluster node. The output from the **lsgroup** command for all nodes is displayed on the node on which the command was executed.

If you specify a group name that does *not* exist on a cluster node, the **cl_lsgroup** command displays a warning message but continues execution of the command on all of the other cluster nodes.

To list all the groups defined on each cluster node using the C-SPOC utility SMIT interface:

- Enter the fastpath `smit cl_lsgroup`

  SMIT displays the following command status window.

```
                        COMMAND STATUS

Command: OK            stdout: yes           stderr: no

Before command completion, additional instructions may appear below.

[TOP]
cav      system 0       true    root
cav      staff  1       false   daemon
cav      bin    2       true    root,bin
cav      sys    3       true    root,bin,sys
cav      adm    4       true    bin,adm
cav      uucp   5       true    nuucp,uucp
```

```
cav      mail   6       true
cav      security       7       true    root
cav      cron   8       true    root
cav      printq 9       true
cav      audit  10      true    root
cav      ecs    28      true
cav      nobody -2      false   nobody,lpd
[MORE...56]
```

Adding Groups on Cluster Nodes

To define a new group on AIX 5L systems, you use the **mkgroup** command. This command adds an entry for the new group to various system security files, including **/etc/group** and **/etc/security/group.** For more information about the **mkgroup** command, see its man page.

You can also define a new group on all cluster nodes from C-SPOC as described in the following procedure. The C-SPOC command performs some verification and then calls the AIX 5L **mkgroup** command on each cluster node to create the group you specify.

If a group with the same name already exists on a cluster node, the operation is aborted. By default, the C-SPOC command requires that the nodes in the HACMP cluster must be powered up and accessible over the network; otherwise, the command fails with an error.

To define a new AIX 5L group on cluster nodes using the C-SPOC utility:

1. Enter the fastpath `smit cl_mkgroup`

   SMIT displays the **Add a Group** panel.
2. Enter data in entry fields to create the group account. The **Group Name** is the only required field. Note, however, that you should also specify the **Group ID**.
3. After you finish filling in the SMIT fields, press Enter. The C-SPOC command executes, creating the new group on all cluster nodes.

Changing Characteristics of Groups in a Cluster

In AIX 5L, you can change the attributes of a group by using either:

- The **chgroup** command
- The AIX 5L SMIT **Change Group Attributes** panel.

For more information about the **chgroup** command, see its man page.

The **chgroup** command modifies the user information stored in the **/etc/group** and the **/etc/security/group** files.

You can also change the attributes of a group on all cluster nodes from C-SPOC as described in the following procedure. This procedure executes the AIX 5L **chgroup** command on each cluster node.

Changing group characteristics from C-SPOC requires that:

- All cluster nodes are accessible
- The Cluster Communications daemon is running
- A group with the name specified exists on all cluster nodes.

Optionally, you can force the C-SPOC command to continue processing even if it encounters an error on one of the cluster nodes.

To change the attributes of a group on all cluster nodes using the C-SPOC utility:

1. Enter the fastpath `smit cl_chgroup`

   SMIT displays the **Change a Group** panel.

2. Specify the name of the group you want to change and press Enter.

   Press F4 to obtain a listing of groups from which to select. SMIT displays a complete listing of the attributes of the group specified, with their current values filled in.

3. Change the value of any group attribute and press Enter.

   The command executes, writing the new attribute value in the appropriate system security files on all cluster nodes.

Removing Groups from the Cluster

To delete a group on an AIX 5L system, you use the **rmgroup** command. This command removes the entry for the group from the **/etc/group** and **/etc/security/group** files. Users that are members of the group are *not* deleted.

If the group is the primary group for any user, the remove operation fails unless you redefine the user's primary group with the **chuser** command. (For more information about using the **chuser** command, see the section Managing Group Accounts.) Only the root user can remove an administrative group or a group with administrative users as members.

To remove a group from all cluster nodes, complete the steps in the following procedure. C-SPOC performs some cluster-wide verification checks and then calls the AIX 5L **rmgroup** command to remove the group on each cluster node.

If a group with the name specified does *not* exist on one of the cluster nodes, the command reports a warning message but continues the operation on the other cluster nodes. By default, the command requires that all cluster nodes are powered up and accessible over the network; otherwise, the command fails with an error. Optionally, you can force the command to continue processing even if it encounters an error on one of the cluster nodes.

To remove a group from cluster nodes using the C-SPOC utility:

1. Enter `smit cl_rmgroup`

   SMIT displays the **Remove a Group** panel.

2. Enter the name of the group you want to remove. Press the F4 key to get a listing of available groups from which to select. After specifying the group name, press Enter.

   The command executes, removing the group from all cluster nodes.

Chapter 17: Managing Cluster Security

This chapter describes how to configure security options to protect your HACMP cluster. The chapter includes the following sections:

- Overview
- Configuring Cluster Security
- Configuring Connection Authentication
- Setting Up Cluster Communications over a VPN
- Configuring Message Authentication and Encryption.

Overview

You can protect access to your HACMP cluster by setting up security for cluster communications between nodes. HACMP provides security for connections between nodes, with higher levels of security for inter-node communications provided through Kerberos (on SP node only) or through virtual private networks. In addition, you can configure authentication and encryption of the messages sent between nodes.

Configuring Cluster Security

HACMP secures communications between cluster nodes for HACMP operations by providing:

- Connection authentication for each new connection request

  On cluster that use SP nodes, you can also use Kerberos for connection authentication.

  Connection authentication can be configured to work in conjunction with virtual private networks (VPNs).
- (*Optional*) Message authentication

  Messages are signed on the sending node, and this signature is verified on the receiving node.
- (*Optional*) Message encryption

Messages are encrypted on the sending node and decrypted on the receiving node, using a common, shared (symmetric) key.

A Cluster Communications daemon (**clcomd**) runs on each HACMP node to transparently manage inter-node communications for HACMP. This daemon consolidates communication mechanisms in HACMP and decreases management traffic on the network. This communication infrastructure requires only one common communication path, rather than multiple TCP connections, between each pair of nodes.

The Cluster Communications daemon logs information about all attempted connections (those accepted and those rejected) to **clcomd.log**. For more information about **clcomd.log**, see Chapter 10: Monitoring an HACMP Cluster.

Although most components communicate through the Cluster Communications daemon, the following HACMP components use a different mechanism for inter-node communications:

Component	Communication Method
Cluster Manager	RSCT
Heartbeat messaging	RSCT
Cluster Information Program (Clinfo)	SNMP

Configuring Connection Authentication

HACMP provides two modes for connection authentication:

- *Standard*. Standard security mode checks source IP address and port values against an access list and uses the principle of *least-privileged*, that is, only the minimum set of access privileges, to perform a specified task for remote command execution. Standard security is the default security mode. See the section Standard Security Mode.
- *Kerberos*. Kerberos security mode provides the features in standard security *and* uses Kerberos security. Kerberos security is available only for the SP. See the section Kerberos Security Mode.

For added security, you can use VPN tunnels between cluster nodes. In this case, traffic for IP interfaces/addresses configured in HACMP is sent through VPN tunnels. See the section Setting Up Cluster Communications over a VPN.

Standard Security Mode

In standard security mode, HACMP authenticates requests for incoming connections by checking the following:

- Source IP address
- Port number
- User privilege

Remote command execution for commands in **/usr/es/sbin/cluster** uses the principle of *least privileged*. This ensures that no arbitrary command can run on a remote node with root privilege. A select set of HACMP commands are considered *trusted* and allowed to run as root; all other commands run as user *nobody*.

Since HACMP 5.1, the dependency on **rsh** and the **~/.rhosts** file to configure host access has been eliminated. Although this file is optional, some commands external to HACMP—for example user-defined event scripts and user programs—may still require an **~/.rhosts** file. HACMP now relies on an internal HACMP trusted host file, **/usr/es/sbin/cluster/etc/rhosts**. to authenticate HACMP communications.

Also, see Understanding the /usr/es/sbin/cluster/etc/rhosts File in Chapter 1: Administering an HACMP Cluster.

Note: The **~/.rhosts** file is required for migration from HACMP or HAES 4.5 to HACMP 5.1 or later. After the migration is complete, the **~/.rhosts** file should be removed if it is *not* required for **rsh** communications by other programs.

To manage inter-node communications, the Cluster Communications daemon requires a list of valid cluster IP labels or addresses to use. There are two ways to provide this information:

- Automatic node configuration
- Individual node configuration (more secure)

Note: During discovery, each node that receives a connect request checks the **/usr/es/sbin/cluster/etc/rhosts** file to ensure that the request is from a legitimate cluster node. The **smit.log** file indicates whether this file is missing or has incorrect entries.

Automatic Configuration of the /usr/es/sbin/cluster/etc/rhosts File

In general, you do *not* need to edit the **/usr/es/sbin/cluster/etc/rhosts** file unless you have specific needs or concerns. HACMP manages connections for you automatically.

If you are configuring HACMP for the first time, the **/usr/es/sbin/cluster/etc/rhosts** file on a node is empty. As a result, the node does *not* have information about the IP labels/addresses for cluster nodes. The first time that a connection is made from another node, such as for HACMP verification, HACMP accepts the connection and adds IP labels/addresses to the **/usr/es/sbin/cluster/etc/rhosts** file. Thereafter HACMP communicates among nodes that have IP labels/addresses listed in the **/usr/es/sbin/cluster/etc/rhosts** file. When you verify and synchronize the cluster, HACMP populates the **~/.rhosts** file on each node with a list of valid IP labels/addresses.

WARNING: If an unwelcome host requests a connection to the node prior to the first connection from another HACMP cluster node, the unwelcome host's connection will be successful, with the host's IP address added to the **/usr/es/sbin/cluster/etc/rhosts** file.

For added security, to make sure that an unwelcome host does *not* connect to a node between the time when you install HACMP software and the time when you initiate a connection from one cluster node to another, you can edit the **/usr/es/sbin/cluster/etc/rhosts** file to add one or more IP labels/addresses (that will be part of your cluster) to the file. For information about editing the **/usr/es/sbin/cluster/etc/rhosts** file, see the section Manually Configuring /usr/es/sbin/cluster/etc/rhosts file on Individual Nodes.

Note: The **clcomd.log** file logs information about all attempted connections. For information about the **clcomd.log** file, see Chapter 2: Using Cluster Log Files in the *Troubleshooting Guide*.

Manually Configuring /usr/es/sbin/cluster/etc/rhosts file on Individual Nodes

For a more secure initial configuration, you can manually configure a **/usr/es/sbin/cluster/etc/rhosts** file for HACMP on each node before configuration. The HACMP installation creates this empty file with read-write permissions for root only. Ensure that each IP address/label is valid for the cluster, otherwise an error is logged in **smit.log** and **clcomd.log**.

To manually set up the **/usr/es/sbin/cluster/etc/rhosts** file:

1. As root, open the file **/usr/es/sbin/cluster/etc/rhosts** on a node.
2. Edit the file to add all possible network interface IP labels or addresses for each node. Put only one IP label or address on each line. Do *not* add any other characters or comments.The format of this file does *not* allow to have comments, additional lines, or characters in it, besides the IP labels.

Disabling and Enabling the Cluster Communications Daemon

If you want to remove the reliance on the Cluster Communications daemon you can turn off the Cluster Communications daemon or rename the **usr/es/sbin/cluster/etc/rhosts** file.

WARNING: If you disable the Cluster Communications daemon or delete the **/usr/es/sbin/cluster/etc/rhosts** file, programs that require inter-node communication, such as C-SPOC, cluster verification and synchronization, file collections, and message authentication and encryption will no longer function.

To stop the Cluster Communications daemon, use the following command:

```
stopsrc -s clcomdES
```

If you want to make changes to the configuration for your HACMP cluster after disabling the daemon, you need to restart the Cluster Communications daemon and supply a valid **rhosts** file.

To start the Cluster Communications daemon, use the following command:

```
startsrc -s clcomdES
```

Troubleshooting the Cluster Communications Daemon

In some cases, if you change or remove IP addresses in the AIX 5L adapter configuration, and this takes place *after* the cluster has been synchronized, the Cluster Communications daemon cannot validate these addresses against the **/usr/es/sbin/cluster/etc/rhosts** file or against the entries in the HACMP's Configuration Database, and HACMP issues an error.

Or, you may obtain an error during the cluster synchronization.

In this case, you must update the information that is saved in the **/usr/es/sbin/cluster/etc/rhosts** file on all cluster nodes, and refresh **clcomd** to make it aware of the changes. When you synchronize and verify the cluster again, **clcomd** starts using IP addresses added to HACMP Configuration Database.

To refresh the Cluster Communications daemon, use:

```
refresh -s clcomdES
```

Also, configure the **/usr/es/sbin/cluster/etc/rhosts** file to contain all the addresses currently used by HACMP for inter-node communication, and then copy this file to all cluster nodes.

Kerberos Security Mode

Kerberos is a network authentication protocol used on the SP. Based on a secret key encryption scheme, Kerberos offers a secure authentication mechanism for client/server applications. It centralizes command authority by using one authentication server, normally configured to be the SP control workstation.

For a more detailed explanation of Kerberos and the security features of the SP system, refer to the *IBM Parallel System Support Programs for AIX 5L Administration Guide.*

In addition, PSSP 3.4 and greater provides the option of running an RS/6000 SP system with an enhanced level of security, and you can use DCE authentication rather than Kerberos 4 authentication. However, these options may affect your HACMP functionality. Read the following sections before planning your cluster security.

Configuring Kerberos Security with HACMP for AIX

By setting up all network IP labels in your HACMP configuration to use Kerberos authentication, you reduce the possibility of a single point of failure. You can configure Kerberos for a cluster automatically by running a setup utility called **cl_setup_kerberos**. Alternatively, you can perform the process manually. Because the utility-based approach is faster and less prone to error, it is usually preferable to the manual method.

To configure Kerberos on the SPs within an HACMP cluster, perform these general steps (where needed detailed procedures appear in the following sections):

Step	What you do
1	Make sure that HACMP has been properly installed on all nodes in the cluster. For information about installing HACMP, see the *Installation Guide*.
2	Configure HACMP cluster topology on one node in the cluster. Note that because the **cl_setup_kerberos** utility needs an initial Kerberized **rcmd** path to each node in the cluster and to the control workstation, you must include the SP Ethernet as part of the configuration. Note that the **setup_authent** script is usually used to configure Kerberos on the entire SP system. The **setup_authent** script creates **rcmd** (used for **rsh** and **rcp**) service principals for all network IP labels listed in the System Data Repository (SDR). The SDR does *not* allow multiple IP labels to be defined on the same interface. However, HACMP requires that multiple IP labels be defined for the same interface during IPAT configurations. For these reasons, each time the nodes are customized after the SP **setup_authent** script is run (through **setup_server** or alone), you must rerun the **cl_setup_kerberos** script or manually reconfigure the systems to use Kerberos.
3	Create new Kerberos service principals and configure all IP labels for Kerberos authentication. You can choose to perform these tasks in either of the following ways: • Automatically—see the section Configuring Kerberos Automatically • Manually—see the section Configuring Kerberos Manually.
4	Set the cluster security mode to **Kerberos**, and then synchronize the cluster.
5	Delete (or at least edit) the **cl_krb_service** file, which contains the Kerberos service principals password you entered during the configuration process. At the very least, you should edit this file to prevent unauthorized users from obtaining the password and possibly changing the service principals.
6	Consider removing unnecessary ~/**.rhosts** files. HACMP does *not* require the traditional TCP/IP access control lists provided by these files (but other applications might). Consult your cluster administrator before removing any version of this file.

Configuring Kerberos Automatically

The **cl_setup_kerberos** utility performs the following tasks:

- Creates new Kerberos service principals in the Kerberos Authentication Database by copying the IP labels from the **cl_krb_service** file
- Extracts the service principals and places them in a new Kerberos services file, **cl_krb-srvtab**
- Creates a **cl_klogin** file that contains additional entries required by the **.klogin** file
- Updates the **.klogin** file on the control workstation and on all nodes in the cluster
- Concatenates the **cl_krb-srvtab** file to each node's **/etc/krb-srvtab** file.

Note: Make sure that you have already installed HACMP on at least one node, and that you have configured cluster topology before performing the following procedure.

To run the **cl_setup_kerberos** utility:

1. Verify that there is a valid **/.k** file on the control workstation. This file stores the Kerberos authentication password so that batched commands can be run. If the **/.k** file is *not* present, issue the following command locally on the control workstation:

   ```
   /usr/lpp/ssp/kerberos/etc/kstash
   ```

2. Run **cl_setup_kerberos** from the configured node. (This utility is found in the **/usr/es/sbin/cluster/sbin** directory.)

   Note: You must be *within* the directory to run this command successfully. It is *not* sufficient to define the path correctly; the only way to run the **cl_setup_kerberos** command correctly is from within the **/usr/es/sbin/cluster/sbin** directory.

   The **cl_setup_kerberos** utility extracts the HACMP IP labels from the configured node and creates a file, **cl_krb_service**, that contains all of the IP labels and additional format information required by the **add_principal** Kerberos utility. It also creates the **cl_adapters** file that contains a list of the IP labels required to extract the service principals from the authentication database.

3. When prompted, enter a Kerberos password for the new principals:

   ```
   Password:
   ```

   Note: The password is added to the **cl_krb_service** file. This password can be the same as the Kerberos Administration Password, but it does *not* have to be. Follow your site's password security procedures.

Configuring Kerberos Manually

To properly configure Kerberos on all HACMP-configured networks, perform the following steps:

Step	What you do
1	Add an entry for each new Kerberos service principal to the Kerberos Authentication Database. The **rcmd.localhost** principal is also required. Verify that it is already present and if it is *not*, manually add this principal. See the section Adding New Service Principals to the Authentication Database.
2	Update the **krb-srvtab** file by extracting each newly added instance from the Kerberos Authentication Database. See the section Updating the krb-srvtab File.
3	Add the new service principals to each node's **/.klogin** file. See the section Adding Kerberos Principals to Each Node's .klogin File.
4	Add the new service principals to each node's **/etc/krb.realms** file. See the section Adding Kerberos Principals to Each Node's /etc/krb.realms File.

Adding New Service Principals to the Authentication Database

To add new service principals to the Kerberos Authentication Database for each network interface:

1. On the control workstation, start the **kadmin** utility:

   ```
   kadmin
   ```

   A welcome message appears.

2. At the `admin:` prompt, type the **add_new_key** command with the name and instance of the new principal:

   ```
   admin: ank service_name.instance
   ```

   where *service_name* is the service (**rcmd**) and *instance* is the address label to be associated with the service. For example, using the service **rcmd** and address label **i1_sw,** the command is:

   ```
   admin: ank rcmd.i1_sw
   ```

3. When prompted, enter the Kerberos Administration Password.

   ```
   Admin password: password
   ```

4. When prompted, enter a Kerberos password for the new principal.

   ```
   Password for service_name.instance: password
   ```

   Note: The password can be the same as the Kerberos Administration Password, but does *not* have to be. Follow your site's password security procedures.

5. Verify that you have added the new principals to the Kerberos database:

   ```
   kdb_util dump /tmp/testdb
   cat /tmp/testdb
   ```

   Remove this copy of the database when you have finished examining it:

   ```
   rm /tmp/testdb
   ```

Updating the krb-srvtab File

To update the **krb-srvtab** file and propagate new service principals to the HACMP cluster nodes:

1. Extract each new service principal for each instance you added to the Kerberos Authentication Database for those nodes you want to update. (This operation creates a new file in the current directory for each instance extracted.)

   ```
   usr/lpp/ssp/kerberos/etc/ext_srvtab -n i1_sw i1_en i1_tr
   ```

2. Combine these new files generated by the **ext_srvtab** utility into one file called *node_name-new-srvtab*:

   ```
   cat i1_sw-new-srvtab i1_en-new-srvtab i1_tr-new-srvtab
   > node_name-new-srvtab
   ```

   The new file appears in the directory where you typed the command.

   Note: Shared labels (used for non-concurrent resource groups with the Online Using Node Distribution Policy startup) need to be included in every **krb-srvtab** file (for nodes in that resource group), so you must concatenate each shared-label **srvtab** file into each *node_name-new-srvtab* file.

3. Copy each *node_name-new-srvtab* file to its respective node.

4. Make a copy of the current **/etc/krb-srvtab** file so that it can be reused later if necessary:

   ```
   cp /etc/krb-srvtab /etc/krb-srvtab-date
   ```

   where *date* is the date you made the copy.

5. Replace the current **krb-srvtab** file with the new *node_name-new-srvtab* file:

   ```
   cp node_name-new-srvtab /etc/krb-srvtab
   ```

6. Verify that the target node recognizes the new principals by issuing the following command on it:

   ```
   ksrvutil list
   ```

   You should see all the new principals for each network interface on that node; if *not*, repeat this procedure.

Adding Kerberos Principals to Each Node's .klogin File

To add the new Kerberos principals to the **/.klogin** file on each HACMP cluster node:

1. Edit the **/.klogin** file on the control workstation to add the principals that were created for each network instance:

   ```
   vi /.klogin
   ```

   Here is an example of the **/.klogin** file for two nodes, i and j. ELVIS_IMP is the name of the realm that will be used to authenticate service requests. Each node has the SP Ethernet, a Token Ring service, and an Ethernet service adapter.

   ```
   root.admin@ELVIS_IMP
   rcmd.i1@ELVIS_IMP
   rcmd.i1_ensvc@ELVIS_IMP
   rcmd.i1_trsvc@ELVIS_IMP
   rcmd.j1@ELVIS_IMP
   rcmd.j1_ensvc@ELVIS_IMP
   rcmd.j1_trsvc@ELVIS_IMP
   ```

2. Copy the **/.klogin** file from the control workstation to each node in the cluster.

To verify that you set this up correctly, issue a Kerberized **rsh** command on all nodes using one of the newly defined interfaces. For example:

```
/usr/lpp/ssp/rcmd/bin/rsh i1_ensvc date
```

To eliminate single points of failure, you should add Kerberos **rcmd** principals for every interface configured in HACMP.

Adding Kerberos Principals to Each Node's /etc/krb.realms File

To add the new Kerberos principals to the **/etc/krb.realms** file on each HACMP cluster node:

1. Edit the **/etc/krb.realms** file on the control workstation and add the principals that were created for each network instance:

   ```
   vi /etc/krb.realms
   ```

   Here is an example of the **krb.realms** file for two nodes, i and j. ELVIS_IMP is the name of the realm that will be used to authenticate service requests. Each node has the SP Ethernet, a Token-Ring service, and an Ethernet service adapter.

   ```
   root.admin ELVIS_IMP
   i1 ELVIS_IMP
   i1_ensvc ELVIS_IMP
   i1_trsvc ELVIS_IMP
   j1 ELVIS_IMP
   j1_ensvc ELVIS_IMP
   j1_trsvc ELVIS_IMP
   i1 ELVIS_IMP
   i1_ensvc ELVIS_IMP
   i1_trsvc ELVIS_IMP
   j1 ELVIS_IMP
   j1_ensvc ELVIS_IMP
   j1_trsvc ELVIS_IMP
   ```

2. Copy the **/etc/krb.realms** file from the control workstation to each node in the cluster.

Setting the HACMP Security Mode

You can set or change the security mode on all nodes in a cluster.

To set the security mode on all nodes in a cluster to use Kerberos:

1. Enter the fastpath `smitty cl_admin`
2. Select **HACMP Security and Users Management** > **Change/Show HACMP Security Mode**.

   The **Change/Show HACMP Security** SMIT panel appears.
3. Set the security mode to **Kerberos**.
4. Synchronize the cluster.

   For more information, see Chapter 7: Verifying and Synchronizing an HACMP Cluster.

Setting Up Cluster Communications over a VPN

You can set up a VPN for HACMP inter-node communications that use the Cluster Communications daemon. In most cases, you set up security when the nodes are *not* publicly available.

Note: For additional security, you can set up your configuration so that RSCT and SNMP also use the VPN tunnel.

VPN support relies on the IP Security feature in AIX 5L. IP Security is separately installable and loadable. The core filesets that need to be installed are:

bos.net.ipsec.rte—The runtime environment for the kernel IP Security environment and commands

bos.msg.LANG.net.ipsec where LANG is the desired language, such as en_US

Also, the following filesets need to be installed for Internet key exchange tunnel support:

bos.net.ipsec.keymgt

bos.net.ipsec.websm

The **bos.crypto** fileset that is appropriate for your country

You use SMIT or the Web-based System Manager to manage a VPN. Refer to your VPN documentation for information about setting up a VPN. For more information about VPNs, go to the following URL:

http://www.ibm.com/servers/aix/products/ibmsw/security/vpn/techref/

To set up HACMP to use a VPN connection for inter-node communications through the Cluster Communications daemon:

1. Configure IP labels on each node.
2. Configure persistent IP labels on each node for use by the VPN.

   For information about configuring persistent IP labels, see Chapter 3: Configuring an HACMP Cluster (Standard).

3. Synchronize the configuration.

   For information about synchronizing configuration, see Chapter 7: Verifying and Synchronizing an HACMP Cluster.

   Note: If the configuration for persistent IP labels is *not* synchronized and you configure HACMP to use persistent labels for VPN tunnels, the cluster nodes will be unable to communicate with each other.

4. *On one node*, configure security to use the persistent labels:

   a. Enter `smit hacmp`

   b. In SMIT, select **Enhanced Configuration > Security and Users Configuration > HACMP Cluster Security > Configure Connection Authentication Mode** and press Enter.

   c. For **Use Persistent Labels for VPN Tunnels**, select **Yes**.

   d. Synchronize the configuration.

5. In the VPN configuration for the VPN implementation that you are using, specify the port to be used for the tunnel.

   By default, the port to be tunneled is 6191. To check whether the port number uses the default value, or another one, review the values in the **/etc/services** file.

6. It is recommended that you close the port used by the VPN tunnel on IP labels other than the persistent IP label being used by the VPN.

   You can close these tunnels from AIX 5L:

   a. Enter `smitty tcpip`

   b. In SMIT, select **Configure IP Security**, then select the appropriate option.

Configuring Message Authentication and Encryption

In addition to connection authentication, you can secure the messages sent through the Cluster Communications Daemon between cluster nodes by authenticating and encrypting those messages. You can use message encryption with message authentication, but you cannot use *not* message encryption alone. Message authentication and encryption are disabled by default. For information about components that do *not* use the Cluster Communications Daemon, see the section Configuring Cluster Security.

Both message authentication and message encryption rely on *secret key* technology. For authentication, the message is signed and the signature is encrypted by a key when sent, and the signature is decrypted and verified when received. For encryption, the encryption algorithm uses the key to make data unreadable. The message is encrypted when sent and decrypted when received.

Message authentication and encryption rely on Cluster Security (CtSec) Services in AIX 5L, and use the encryption keys available from Cluster Security Services. HACMP message authentication uses message digest version 5 (MD5) to create the digital signatures for the message digest. Message authentication uses the following types of keys to encrypt and decrypt signatures and messages (if selected):

- Data encryption standard (DES)
- Triple DES
- Advanced encryption standard (AES).

The message authentication mode is based on the encryption algorithm. Your selection of a mesage authentication mode depends on the security requirements for your HACMP cluster.

For information about network security for the AIX 5L operating system, see the following URL:

http://publib16.boulder.ibm.com/doc_link/en_US/a_doc_lib/aixbman/security/securitytfrm.htm

Authenticating and encrypting messages increases the overhead required to process messages and may impact HACMP performance. Processing more sophisticated encryption algorithms may take more time than less complex algorithms. For example, processing AES messages may take more time than processing DES messages.

Prerequisites

The HACMP product does *not* include encryption libraries. Before you can use message authentication and encryption, the following AIX 5L filesets must be installed on each cluster node:

- For data encryption with DES message authentication: **rsct.crypt.des**
- For data encryption standard Triple DES message authentication: **rsct.crypt.3des**
- For data encryption with Advanced Encryption Standard (AES) message authentication: **rsct.crypt.aes256**

You can install these filesets from the AIX 5L Expansion Pack CD-ROM.

If you install the AIX 5L encryption filesets after you have HACMP running, restart the Cluster Communications daemon to enable HACMP to use these filesets. To restart the Cluster Communications daemon:

```
stopscr -s clcomdes
startsrc -s clcomdes
```

If your configuration includes persistent labels, make sure that this configuration is synchronized before proceeding.

WARNING: Do *not* perform other cluster configuration activities while you are configuring message authentication and encryption for a cluster. Doing so may cause communication problems between nodes. Make sure that security configuration is complete and the cluster synchronized before performing other configuration tasks.

Managing Keys

HACMP cluster security uses a shared common (symmetric) key. This means that each node must have a copy of the *same* key for inter-node communications to be successful. You control when keys change and how keys are distributed.

You can allow HACMP to distribute a key for you, or you can manually copy a key to each node in a cluster. Copying a key to each cluster node can provide a higher level of security than having HACMP distribute the key, depending on the method you use to copy the key to the cluster nodes.

HACMP key distribution (disabled by default) is only as secure as the network over which a key is distributed. When HACMP distributes a key to other nodes, it requires confirmation that the intent is to have HACMP distribute keys. Confirmation information appears in the **clcomd.log** file.

WARNING: If the key is intercepted during distribution by an unwelcome party, the security of the system can be compromised.

Cluster synchronization does *not* update keys and does *not* distribute keys among nodes.

Location of Keys

On each node, a key is stored in the **/usr/es/sbin/cluster/etc** directory. The name of the key identifies the encryption type selected:

- key_md5_des
- key_md5_3des
- key_md5_aes

When to Generate and Distribute a Key

Generate and distribute a key after:

- Enabling message authentication
- Changing the configuration for message authentication.

Also, change the key in accordance with the security policies for your organization.

Note: Communication between cluster nodes requires that all nodes have active copies of the same key. You activate a new key after you distribute the key to each node in the cluster.

If you change the configuration, you may have two different keys in the **/usr/es/sbin/cluster/etc** directory, the active one for your configuration and the one from the previous configuration.

Note: To prevent a configuration error, delete the key that is no longer being used from each cluster node.

About Configuring Message Authentication and Encryption

How you configure message authentication and encryption depends on the method you use to distribute a key: automatically through HACMP or manually by coping a key to each cluster node.

Ensure that the message authentication and encryption configuration is consistent across cluster nodes; otherwise, HACMP cannot communicate between cluster nodes.

Configuring Message Authentication and Encryption using Automatic Key Distribution

Make sure that the cluster is synchronized before you start to configure message authentication and encryption. This ensures that cluster nodes can communicate with each other.

Step 1: Enable Automatic Key Distribution on Each Node

To make sure that you can distribute a new key through HACMP, enable **Automatic Key Distribution** *on each node* in the cluster *before:*

- You change the message authentication mode
- You try to automatically distribute a key to cluster nodes.

To enable key distribution on each cluster node:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Security and Users Configuration > HACMP Cluster Security > Configure Message Authentication Mode and Key Management > Enable/Disable Automatic Key Distribution** and press Enter.

   The **Enable/Disable Automatic Key Distribution** panel appears.
3. For **Enable Key Distribution**, select **Yes**.
4. Repeat step 1.through step 3. on the other nodes in the cluster.

Step 2: Enable or Change Message Authentication

You enable or change message authentication and encryption *from one cluster node*.

To enable or change message authentication:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Security and Users Configuration > HACMP Cluster Security > Configure Message Authentication Mode and Key Management > Configure Message Authentication Mode** and press Enter.

   The **Configure Message Authentication Mode** panel appears.
3. Enter field values as follows:

Message Authentication Mode	Select one of the following modes: **MD5_DES** The MD5 algorithm is used for message digest (signature) and the DES algorithm is used for signature encryption. **MD5_3DES** The MD5 algorithm is used for message digest (signature) and the triple DES algorithm is used for signature encryption. **MD5_AES** The MD5 algorithm is used for message digest (signature) and the AES algorithm is used for signature encryption. **None** This indicates that neither message authentication nor message encryption is being used.

Enable Encryption	Select **Yes** to *enable* message encryption for messages sent between HACMP nodes. Select **No** to *disable* message encryption for messages sent between HACMP nodes.

4. Press Enter.

Step 3: Generate and Distribute a Key from One Node

If you are enabling or changing message authentication and encryption, complete this procedure on the same node where you completed Step 2: Enable or Change Message Authentication.

To generate a new key and distribute it through HACMP:

1. In SMIT, select **Extended Configuration > Security and Users Configuration > HACMP Cluster Security > Configure Message Authentication Mode and Key Management > Generate/Distribute a Key** and press Enter.

   The **Generate/Distribute a Key** panel appears.

2. Enter field values as follows:

Type of Key to Generate	Lists the active authentication mode
Distribute a Key	**Yes**

3. When prompted, confirm that you want HACMP to distribute a key. This information is written to the **/var/hacmp/clcomd.clcomd.log** file.

   Note: If for some reason SMIT cannot copy the key to cluster nodes, copy the key file to diskette and copy it to the node. For information about how to manually distribute a key, see the section Step 2: Distribute a New Key by Copying It to Cluster Nodes in the section Configuring Message Authentication and Encryption using Manual Key Distribution.

Step 4: Activate the Key on Each Node

After you distribute a new key to each node in the cluster, on the node from which you distributed the key, active it for *all cluster node*s. This action makes it possible for cluster nodes to communicate with each other.

To activate a new key:

1. In SMIT, select **Extended Configuration > Security and Users Configuration > HACMP Cluster Security > Configure Message Authentication Mode and Key Management > Activate the New Key on All Cluster Nodes** and press Enter.

   SMIT displays `Are you sure?`

2. Press Enter to activate the key on all cluster nodes.

   The **Command Status** panel lists the nodes on which the key is active.

Step 5: Synchronize the Cluster

Synchronize the cluster configuration. For information about synchronizing the cluster, see Chapter 7: Verifying and Synchronizing an HACMP Cluster.

Step 6: Disable Automatic Key Distribution on Each Node

After you distribute a key to cluster nodes through HACMP and activate the key, disable **Automatic Key Distribution** *on each node* in the cluster.

WARNING: Do *not* leave **Automatic Key Distribution** enabled. Doing so might allow an unwelcome user to distribute a spurious key to cluster nodes, which would compromise cluster security.

To disable automation key distribution from each cluster node:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Security and Users Configuration > HACMP Cluster Security > Configure Message Authentication Mode and Key Management > Enable/Disable Automatic Key Distribution** and press Enter.

   The **Enable/Disable Automatic Key Distribution** panel appears.
3. For **Enable Key Distribution**, select **No**.

Configuring Message Authentication and Encryption using Manual Key Distribution

Synchronize the cluster before you start to configure message authentication and encryption. This ensures that cluster nodes can communicate with each other.

Step 1: Enable or Change Message Authentication and Encryption

You enable message authentication and encryption from one cluster node.

To enable or change message authentication:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Security and Users Configuration > HACMP Cluster Configuration > Configure Message Authentication Mode and Key Management > Configure Message Authentication Mode** and press Enter.

   The **Configure Message Authentication Mode** panel appears.

3. Enter field values as follows:

Message Authentication Mode	Select one of the following modes: **MD5_DES** The MD5 algorithm is used for message digest (signature) and the DES algorithm is used for signature encryption. **MD5_3DES** The MD5 algorithm is used for message digest (signature) and the triple DES algorithm is used for signature encryption. **MD5_AES** The MD5 algorithm is used for message digest (signature) and the AES algorithm is used for signature encryption. **None** This indicates that neither message authentication nor message encryption is being used.
Enable Encryption	Select **Yes** to *enable* message encryption for messages sent between HACMP nodes. Select **No** to *disable* message encryption for messages sent between HACMP nodes.

4. Press Enter.

Step 2: Distribute a New Key by Copying It to Cluster Nodes

Ensure that you distribute the same encryption key to each cluster node; otherwise, HACMP cannot communicate between cluster nodes.

To generate a new key and copy it to other cluster nodes:

1. On the node where you want to create a key, enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Security and Users Configuration > HACMP Cluster Security > Configure Message Authentication Mode and Key Management > Generate/Distribute a Key** and press Enter.

   The **Generate/Distribute a Key** panel appears.
3. Enter field values as follows:

Type of Key to Generate	Lists the active authentication mode
Distribute a Key	**No**

4. Copy the key file from the node where the key was generated to each node in the HACMP cluster.

   On each node, a key is stored in the **/usr/es/sbin/cluster/etc** directory. The name of the key identifies the encryption type selected:

   - key_md5_des
   - key_md5_3des
   - key_md5_aes

You can copy the file to diskette and then go to each node and copy the key file to the appropriate directory, or you can use a remote copy command such as **ftp** or **rcp**.

Note: If for some reason you cannot copy the key from one node to another using a command such as **ftp** or **rcp**, then copy the key file to diskette and copy it to the node.

WARNING: A key may already be present on each node, make sure that you copy the key to each node. The new key overwrites the older one if the keys are of the same type, for example if the key is for 3DES. *If the keys on the nodes do not match, HACMP does not function.*

Step 3: Activate the Key on Each Node

After you distribute a new key to each node in the cluster, from one node you activate the key *on all cluster node*s to make it possible for cluster nodes to communicate with each other. If you enabled or changed the message authentication mode, you should activate the key from the cluster node where you made that configuration change.

To activate a new key:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Security and Users Configuration > HACMP Cluster Security > Configure Message Authentication Mode and Key Management > Activate the New Key on All Cluster Nodes** and press Enter.

   SMIT displays `Are you sure?`
3. Press Enter to activate the key on all cluster nodes.

   The **Command Status** panel lists the nodes on which the key is active.

Step 4: Synchronize the Cluster

Synchronize the cluster configuration. For information about synchronizing the cluster, see Chapter 7: Verifying and Synchronizing an HACMP Cluster.

Changing the Security Authentication Mode

The procedure for changing the message authentication mode is the same as the procedure for initially setting up message authentication and encryption for a cluster. See the previous sections:

- Configuring Message Authentication and Encryption using Automatic Key Distribution
- Configuring Message Authentication and Encryption using Manual Key Distribution

Changing a Key

If you want to change a security key, but *not* change the message authentication mode:

To distribute the key through HACMP	Follow the instructions in the section: Configuring Message Authentication and Encryption using Automatic Key Distribution and omit the following step: Step 2: Enable or Change Message Authentication
To distribute the key by copying it to cluster nodes	Follow the instructions in the section: Configuring Message Authentication and Encryption using Manual Key Distribution and omit the following step: Step 1: Enable or Change Message Authentication and Encryption

Troubleshooting Message Authentication and Encryption

Problems using message authentication and encryption are most likely caused by configuration errors.

To remedy configuration problems:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Security and Users Configuration > HACMP Cluster Configuration > Configure Message Authentication Mode and Key Management > Configure Message Authentication Mode** and press Enter.

   The **Configure Message Authentication Mode** panel appears.
3. In the **Configure Message Authentication Mode** panel, set the **Message Authentication Mode** to **None**.
4. Configure message authentication and encryption again as described in this section.

Chapter 18: Saving and Restoring Cluster Configurations

This chapter explains how to use the cluster snapshot utility to save and restore cluster configurations. The following sections explain the utility:

Overview

The cluster snapshot utility allows you to save to a file a record of all the data that defines a particular cluster configuration. This facility gives you the ability to recreate a particular cluster configuration—a process called applying a snapshot—provided the cluster is configured with the requisite hardware and software to support the configuration.

In addition, a snapshot can provide useful information for troubleshooting cluster problems. Because the snapshots are simple ASCII files that can be sent via e-mail, they can make remote problem determination easier.

Note: You can*not* use the cluster snapshot facility in a cluster concurrently running different versions of HACMP.

By default, HACMP does *not* collect cluster log files when you create the cluster snapshot. Cluster snapshots are used for recording the cluster configuration information, whereas cluster logs only record the operation of the cluster and *not* the configuration information. Skipping the log collection reduces the size of the snapshot and speeds up running the snapshot utility. The size of the cluster snapshot depends on the configuration. For instance, a basic two-node configuration requires roughly 40KB.

Note: You can change the default to collect cluster log files using SMIT if you need logs for problem reporting. This option is available under the SMIT menu **Problem Determination Tools > HACMP Log Viewing and Management**. It is recommended to use this option only if IBM support personnel request logs.

You can also add your own custom snapshot methods to store additional user-specified cluster and system information in your snapshots. The output from these user-defined custom methods is reported along with the conventional snapshot information.

Relationship between the OLPW Cluster Definition File and a Cluster Snapshot

There is more than one way of capturing the cluster configuration:

- The cluster snapshot
- The cluster definition file that you create and edit using the Online Planning Worksheets application.

This section clarifies the relationship between these two utilities and helps you to decide when to use each utility.

The Online Planning Worksheets (OLPW) application allows you to save your cluster configuration data, as does the cluster snapshot. In addition, OLPW allows you to *edit* your configuration data, reading data exported from your current HACMP configuration or data exported from a converted snapshot file. However, the set of data saved from OLPW in the cluster definition file is less comprehensive than a cluster snapshot. For example, OLPW does *not* contain all ODM information.

For information about using the Online Planning Worksheets application, see Chapter 9: Using Online Planning Worksheets in the *Planning Guide*.

To help you decide whether you should use a cluster snapshot or a cluster definition file, see the following table, which lists various scenarios:

Use this type of configuration file:	When you are:
Cluster Snapshot	• Upgrading HACMP to the current version • Troubleshooting HACMP configuration problems.
OLPW Cluster Definition File	• Planning your HACMP cluster configuration • Viewing your HACMP cluster configuration in an easy-to-read format • Editing your HACMP cluster configuration information.
Either type	• Capturing cluster and node configuration information to record the state of your cluster.

Information Saved in a Cluster Snapshot

The primary information saved in a cluster snapshot is the data stored in the HACMP Configuration Database classes (such as HACMPcluster, HACMPnode, HACMPnetwork, HACMPdaemons).This information is used to recreate the cluster configuration when a cluster snapshot is applied to nodes installed with HACMP.

The cluster snapshot does *not* save any user-customized scripts, applications, or other non-HACMP configuration parameters. For example, the names of application servers and the locations of their start and stop scripts are stored in the HACMPserver Configuration Database object class. However, the scripts themselves as well as any applications they may call are *not* saved.

The cluster snapshot also does *not* save any device- or configuration-specific data that is outside the scope of HACMP. For instance, the facility saves the names of shared filesystems and volume groups; however, other details, such as NFS options or LVM mirroring configuration are *not* saved.

If you moved resource groups using the Resource Group Management utility **clRGmove**, once you apply a snapshot, the resource groups return to behaviors specified by their default nodelists.

To investigate a cluster after a snapshot has been applied, run **clRGinfo** to view the locations and states of resource groups.

Note: In HACMP 5.2 and up, you can reset cluster tunable values using the SMIT interface. HACMP creates a cluster snapshot, prior to resetting. After the values have been reset to their defaults, you can apply the snapshot and return to customized cluster settings, if needed. For more information, see Resetting HACMP Tunable Values section in Chapter 1: Troubleshooting HACMP Clusters in the *Troubleshooting Guide*.

Format of a Cluster Snapshot

The cluster snapshot utility stores the data it saves in two separate files created in the directory **/usr/es/sbin/cluster/snapshots**:

ODM Data File (.odm)	This file contains all the data stored in the HACMP Configuration Database object classes for the cluster. This file is given a user-defined basename with the **.odm** file extension. Because the Configuration Database information is largely the same on every cluster node, the cluster snapshot saves the values from only one node.
Cluster State Information File (.info)	This file contains the output from standard AIX 5L and HACMP system management commands. This file is given the same user-defined basename file with the **.info** file extension. Output from custom snapshot methods is appended to this file.

Cluster Snapshot ODM Data File

The cluster snapshot Configuration Database data file is an ASCII text file divided into three delimited sections:

Version section	This section identifies the version of the cluster snapshot. The characters <VER identify the start of this section; the characters </VER identify the end of this section. The version number is set by the cluster snapshot software.
Description section	This section contains user-defined text that describes the cluster snapshot. You can specify up to 255 characters of descriptive text. The characters <DSC identify the start of this section; the characters </DSC identify the end of this section.

ODM data section This section contains the HACMP Configuration Database object classes in generic AIX 5L ODM stanza format. The characters <ODM identify the start of this section; the characters </ODM identify the end of this section.

The following is an excerpt from a sample cluster snapshot Configuration Database data file showing some of the ODM stanzas that are saved.

```
<VER
1.0
</VER

<DSC
My Cluster Snapshot
</DSC

<ODM

HACMPcluster:
   id = 97531
   name = "Breeze1"
   nodename = “mynode”
   sec_level = “Standard”
   last_node_ids = “2,3”
   highest_node_id = 3
   last_network_ids = “3,6”
   highest_network_id = 6
   last_site_ides = “ “
   highest_site_id = 0
   handle = 3
   cluster_version = 5
   reserved1 = 0
   reserved2 = 0
   wlm_subdir = “ “

HACMPnode:
   name = “mynode”
   object = “VERBOSE_LOGGING”
   value = “high”
.
.
.
</ODM
```

clconvert_snapshot Utility

You can run **clconvert_snapshot** to convert cluster snapshots from a release supported for upgrade to a recent HACMP release. The **clconvert_snapshot** is *not* run automatically during installation, and you must always run it from the command line. Each time you run the **clconvert_snapshot** command, conversion progress is logged to the **/tmp/clconvert.log** file.

Note: Root user privilege is required to run **clconvert_snapshot**. You must know the HACMP version from which you are converting in order to run this utility.

For more information on the **clconvert_snapshot** utility, refer to the **clconvert_snapshot** man page or to Appendix C: HACMP for AIX Commands.

Defining a Custom Snapshot Method

If you want additional, customized system and cluster information to be appended to the **.info** file, you should define custom snapshot methods to be executed when you create your cluster snapshot.

To define a custom snapshot method, perform the following steps.

1. Enter `smit hacmp`
2. In SMIT, select **HACMP Extended Configuration > Snapshot Configuration > Configure Custom Snapshot Method > Add a Custom Snapshot Method** and press Enter.
3. Enter field values as follows:

Custom Snapshot Method Name	A name for the custom snapshot method you would like to create.
Custom Snapshot Method Description	Add any descriptive information about the custom method.
Custom Snapshot Script Filename	Add the full pathname to the custom snapshot scriptfile.

Once you have defined one or more custom snapshot methods, when you create a cluster snapshot you are asked to specify which custom method(s) you wish to run in addition to the conventional snapshot.

Changing or Removing a Custom Snapshot Method

After you have defined a custom snapshot method, you can change or delete it using the other menu items in the **Configure Custom Snapshot Method** SMIT panel: **Change/Show a Custom Snapshot Method** and **Remove a Custom Snapshot Method**.

When you select one of these menus, a picklist of existing custom snapshot methods appears. Select the one you wish to change or remove and fill in the appropriate fields, or answer the prompt to confirm deletion.

Creating (Adding) a Cluster Snapshot

You can initiate cluster snapshot creation from any cluster node. You can create a cluster snapshot on a running cluster. The cluster snapshot facility retrieves information from each node in the cluster. Accessibility to all nodes is required. The snapshot is stored on the local node.

Note: To get an accurate snapshot of a system that has been configured with Kerberos security, you must set up *all* Kerberos service principals before taking the snapshot. For details about configuring cluster security see Chapter 16: Managing User and Groups.

To create a cluster snapshot:

1. Enter `smit hacmp`
2. In SMIT, select **HACMP Extended Configuration > Snapshot Configuration > Add a Cluster Snapshot** and press Enter.
3. Enter field values as follows:

Cluster Snapshot Name	The name you want for the basename for the cluster snapshot files. The default directory path for storage and retrieval of the snapshot is **/usr/es/sbin/cluster/snapshots**. You can specify an alternate path using the SNAPSHOTPATH environment variable.
Custom Defined Snapshot Methods	Specify one or more custom snapshot methods to be executed if desired; press F4 for a picklist of custom methods on this node. If you select **All**, the custom methods will be executed in alphabetical order on each node.
Save Cluster Log Files in a Snapshot	The default is **No**. If you select **Yes**, HACMP collects cluster log files from all nodes and saves them in the snapshot. Saving log files can significantly increase the size of the snapshot.
Cluster Snapshot Description	Enter any descriptive text you want inserted into the cluster snapshot. You can specify any text string up to 255 characters in length.

Applying a Cluster Snapshot

Applying a cluster snapshot overwrites the data in the existing HACMP Configuration Database classes on all nodes in the cluster with the new Configuration Database data contained in the snapshot. You can apply a cluster snapshot from any cluster node.

Note: Only the information in the **.odm** file is applied. The **.info** file is *not* needed to apply a snapshot.

Applying a cluster snapshot may affect HACMP Configuration Database objects and system files as well as user-defined files.

- If cluster services are inactive on all cluster nodes, applying the snapshot changes the Configuration Database data stored in the system default configuration directory (DCD).
- If cluster services are active on the local node, applying a snapshot triggers a cluster-wide dynamic reconfiguration event.

  During dynamic reconfiguration, in addition to synchronizing the Configuration Database data stored in the DCDs on each node, HACMP replaces the current configuration data stored in the active configuration directory (ACD) with the updated configuration data in the DCD. The snapshot becomes the active configuration. For more information about dynamic reconfiguration of a cluster, see Chapter 15: Managing Resource Groups in a Cluster.

Note: A cluster snapshot used for dynamic reconfiguration may contain changes to the cluster topology and to the cluster resources. You can change both the cluster topology and cluster resources in a single dynamic reconfiguration event.

To apply a cluster snapshot using SMIT, perform the following steps:

1. Enter `smit hacmp`
2. In SMIT, select **HACMP Extended Configuration > Snapshot Configuration > Apply a Cluster Snapshot** and press Enter.

   SMIT displays the **Cluster Snapshot to Apply** panel containing a list of all the cluster snapshots that exist in the directory specified by the SNAPSHOTPATH environment variable.
3. Select the cluster snapshot that you want to apply and press Enter. SMIT displays the **Apply a Cluster Snapshot** panel.
4. Enter field values as follows:

Cluster Snapshot Name	Displays the current basename of the cluster snapshot. This field is *not* editable.
Cluster Snapshot Description	Displays the text stored in the description section of the snapshot files. This field is *not* editable.
Un/Configure Cluster Resources?	If you set this field to **Yes**, HACMP changes the definition of the resource in the Configuration Database and performs any configuration triggered by the resource change. For example, if you remove a filesystem, HACMP removes the filesystem from the Configuration Database and also unmounts the filesystem. By default, this field is set to **Yes.** If you set this field to **No**, HACMP changes the definition of the resource in the Configuration Database but does *not* perform any configuration processing that the change may require. For example, a filesystem would be removed from the HACMP cluster definition but would *not* be unmounted. This processing is left to be performed by HACMP during a fallover. HACMP attempts to limit the impact on the resource group when a component resource is changed. For example, if you add a filesystem to the resource group that already includes the underlying volume group as a resource, HACMP does *not* require any processing of the volume group. Other modifications made to the contents of a resource group may cause the entire resource group to be unconfigured and reconfigured during the dynamic reconfiguration. Cluster clients experience an interruption in related services while the dynamic reconfiguration is in progress.

Force apply if verify fails?	If this field is set to **No,** synchronization aborts if verification of the new configuration fails. As part of dynamic reconfiguration processing, the new configuration is verified before it is made the active configuration. By default, this field is set to **No**. If you want synchronization to proceed even if verification fails, set this value to **Yes**.

Note: In some cases, the verification uncovers errors that do *not* cause the synchronization to fail. HACMP reports the errors in the SMIT command status window so that you are aware of an area of the configuration that may be a problem. You should investigate any error reports, even when they do *not* interfere with the synchronization.

If the apply process fails or you want to go back to the previous configuration for any reason, you can re-apply an automatically saved configuration. See Undoing an Applied Snapshot below for details.

Dynamic Changes and Cluster Snapshots

If you create a cluster snapshot and make a dynamic (DARE) change to a working cluster, such as removing and then re-adding a network, the snapshot may fail due to naming issues. For example, the following steps would make a snapshot fail:

1. Start the cluster.
2. Create a snapshot.
3. Remove a network dynamically.
4. Add a network dynamically using the same name as the one that was removed in step 3.
5. Attempt to apply snapshot from step 2.

However, if you use a *different* network name in step 4 than the network that was removed, you can apply the snapshot successfully. (The problem is that a different network ID is used when the network is added back into the cluster.)

Undoing an Applied Snapshot

Before the new configuration is applied, the cluster snapshot facility automatically saves the current configuration in a snapshot called **~snapshot.*n*.odm**, where **n** is either 1 (the most recent), 2, or 3. The saved snapshots are cycled so that only three generations of snapshots exist. If the apply process fails or you want to go back to the previous configuration for any reason, you can re-apply the saved configuration. The saved snapshot are stored in the directory specified by the SNAPSHOTPATH environment variable.

Changing a Cluster Snapshot

After creating a cluster snapshot, you can change the basename assigned to cluster snapshot files and the description contained in these files. Note that you must use the SMIT interface to perform this task.

To change a cluster snapshot, perform the following steps:

1. Enter `smit hacmp`
2. In SMIT, select **HACMP Extended Configuration > Snapshot Configuration > Change/Show a Cluster Snapshot** and press Enter.

   SMIT displays the **Change/Show a Cluster Snapshot** panel with a list of all the cluster snapshots that exist in the directory specified by SNAPSHOTPATH.
3. Select the cluster snapshot to change and press Enter.
4. Enter field values as follows:

Cluster Snapshot Name	Displays the current basename of the cluster snapshot.
New Cluster Snapshot Name	Enter the new name you want assigned as the basename of the cluster snapshot files.
Cluster Snapshot Description	SMIT displays the current description. You can edit the text using up to 255 characters.

Removing a Cluster Snapshot

Removing a cluster snapshot deletes both of the ASCII files (**.odm** and **.info**) that define the snapshot from the snapshots directory. (The directory in which the snapshots are stored is defined in the SNAPSHOTPATH environment variable.) You must use SMIT to remove a cluster snapshot.

To remove a cluster snapshot using the SMIT interface, perform the following steps:

1. Enter `smit hacmp`
2. In SMIT, select **HACMP Extended Configuration > HACMP Snapshot Configuration > Remove a Cluster Snapshot** and press Enter.

   SMIT generates and displays a list of all the cluster snapshots that exist in the directory specified by the SNAPSHOTPATH environment variable.
3. Select the cluster snapshot to remove and press Enter.

   The cluster snapshot facility deletes the files in the snapshot directory that are associated with that snapshot.

Appendix A: 7x24 Maintenance

The goal of high availability is to keep systems up and running, allowing continuous access to critical applications. In many enterprises it has become necessary to keep applications running seven days a week, 24 hours a day. With proper planning, customizing, and monitoring, an HACMP cluster can provide nearly continuous availability, interrupted only by scheduled, necessary maintenance.

This appendix is a collection of information describing the issues and procedures involved in keeping a cluster running on as close to a 7 X 24 basis as possible.

The appendix contains the following sections:

- Planning for 7 X 24 Maintenance. This section reemphasizes the importance of careful planning and customization of the original installation of the cluster.
- Runtime Maintenance. This section offers reminders and tips to help you avoid actions that endanger a stable, running cluster.
- Hardware Maintenance. This section contains procedures for changing or replacing certain hardware.
- Preventive Maintenance. This section reviews tools you can use to avoid problems or catch them early.

Overview

Throughout all stages of cluster administration—planning, configuration, maintenance, troubleshooting, and upgrading—here are tasks you can do and systems you can put in place to help ensure your cluster's nearly continuous availability.

Once you have configured the cluster and brought it online, it is very important to do maintenance tasks in as non-disruptive a way as possible. The HACMP cluster is a distributed operating system environment. Therefore maintaining an HACMP cluster requires attention to some issues that have different ramifications in the cluster environment compared to maintaining a single-server system.

Making changes to a cluster must be thoroughly planned, since changes to one component may have cascading effects. Changes on one node may affect other nodes, but this may *not* be apparent until fallover occurs (or cannot occur due to a non-synchronized change to the cluster). Some of the do's and don'ts of cluster maintenance are explained in this appendix.

Setting up and following regular preventive maintenance procedures helps alert you to any potential problems before they occur. Then you can take timely action or plan fallovers or cluster downtime at your convenience as necessary to deal with any impending issues.

Planning for 7 X 24 Maintenance

Carefully planning the original installation of your cluster goes a long way toward making cluster maintenance easier. A well-configured and customized cluster is the first step to good preventive maintenance. Proper cluster configuration also makes it less likely you will have to make changes that affect cluster performance while users are accessing their applications.

Planning the cluster starts with a single point of failure analysis. See the *Planning Guide* for a detailed list of issues to consider. Once the cluster is installed and running, you need to handle any failures as quickly and as automatically as possible. Planning for runtime failure recovery helps ensure that HACMP for AIX 5L does all that it is capable of doing to keep your critical resources online.

This section includes information on the following topics:

- Customizing the cluster, including setting up error notification to improve monitoring and management of the cluster
- Tuning the communications system—network and nameserving issues
- Planning disk and volume group layout
- General planning for hardware and software maintenance.

Customizing the Cluster

Customizing the cluster enhances your ability to monitor the cluster and keep it running. You can define a pre-event, a post-event, and a notification method for every cluster event. Notification of events is critical to maintain service for any HACMP cluster. Although HACMP writes messages to the **hacmp.out** and **cluster.log** log files, it is very useful to include notifications to the console or mail to the system administrator when an event occurs that demands immediate attention

You can include automatic recovery actions as well as notification in the cluster customization. Use the HACMP and AIX 5L tools available to customize some or all of the following:

- Hardware error notification
- Hardware failure notification
- Cluster event notification
- Pre- and post-event recovery actions
- Network failure escalation
- ARP cache refresh
- Pager notification
- Application server scripts.

It is highly recommended that you maintain a test cluster as well as your production cluster. Thus before you make any major change to the production cluster, you can test the procedure on the test cluster. HACMP supplies event emulation utilities to aid in testing.

Customizing AIX 5L Error Notification of Hardware Errors

Customizing notification when you configure the cluster is a good preventive measure. See Chapter 7 in the *Planning Guide* for complete information on customizing and setting up notification of cluster events.

See Chapter 9: Configuring AIX 5L for HACMP in the *Installation Guide* for information on using AIX 5L Error Notification, and for information on setting up automatic notification for hardware errors that do *not* cause cluster events.

Using the HACMP Automatic Error Notification SMIT panels, you can turn on automatic error notification for selected hard, non-recoverable error types: disk, disk adapter, and SP switch adapter errors. All disks defined as HACMP resources, and disks in the rootvg and HACMP volume groups and filesystems are included.

You may want to set up error notification for certain media or temporary errors. You may also want to customize the error notification for some devices rather than using one of the two automatic error notification methods.

List of Hardware Errors to Monitor

The following list of hardware errors gives you a good idea of types of errors to monitor. The first list shows which errors are handled by the HACMP automatic error notification utility. The following lists show other types of errors you may want to address. For each device monitored, you can determine an additional action other than notification, such as:

- Stop cluster services and move resource groups to another node.
- Initiate a custom recovery action such as reconfiguration for a failed device using an alternative device.

Hardware Errors Handled by HACMP Auto-Error Notification	
DISK_ERR2	Permanent physical disk error (known error)
DISK_ERR3	Permanent physical disk error, adapter detected (known error)
SCSI_ERR1	Permanent SCSI adapter hardware error (known error)
SCSI_ERR3	Permanent SCSI adapter microcode error (known error)
SCSI_ERR5	Temporary SCSI bus error
SCSI_ERR7	Permanent unknown system error
SCSI_ERR9	Potential Data loss condition
SDA_ERR1	Adapter hardware error condition
SDA_ERR3	Permanent unknown system error
SDC_ERR1	Controller/DASD link error
SDC_ERR2	Controller hardware error
SSA_HDW_ERR	SSA hardware error condition
SSA_DISK_ERR1	Permananent microcode program error

SSA_DISK_ERR4	Permanent disk operation error
DISK_ARRAY_ERR2	Permanent disk operation error (disk failure)
DISK_ARRAY_ERR3	Permanent disk operation error (disk failure)
DISK_ARRAY_ERR5	Permanent disk operation error (disk failure)
SCSI_ARRAY_ERR2	SCSI hardware error
HPS_FAULT4_ER	SP Switch error

Disk and Adapter Errors Not Covered by HACMP Auto-Error Notification	
SSA_HDW_RECOVERED	Temporary adapter error
SSA_DISK_ERR3	Temporary disk operation error
SSA_DEGRADED_ERROR	Adapter performance degraded
SSA_LOGGING _ERROR	Permanent: unable to log an error against a disk
SSA_LINK_OPEN	Permanent adapter detected open serial link
SSA_SOFTWARE_ERROR	Permanent software program error
SSA_LINK_ERROR	Temporary link error
SSA_DETECTED_ERROR	Permanent loss of redundant power/cooling
LVM_MISSPVADDED	PV defined as missing (unknown error)
LVM_SA_WRT	PV defined as missing (unknown error)
LVM_SA_PVMISS	Failed to write VGSA (unknown error)

Disk Array Errors Not Covered by HACMP Auto-Error Notification	
DISK_ARRAY_ERR4	Temporary disk operation error (disk media failing)
DISK_ARRAY_ERR6	Permanent array subsystem degradation (disk media failure)
DISK_ARRAY_ERR7	Permanent array subsystem degradation (controller)
DISK_ARRAY_ERR8	Permanent array active controller switch (controller)
DISK_ARRAY_ERR9	Permanent array controller switch failure

Failed 64-port Adapter (tty device driver)	
COM_PERM_PIO	PIO exception, possible adapter failure

You may have additional devices critical to your operation that are *not* supported by HACMP for AIX 5L. You can set up AIX 5L error notification to monitor microcode errors for those devices or adapter time-outs.

Customizing Cluster Events

Customizing cluster events to send notification or to take recovery actions is another method you can use to help maintain the cluster running as smoothly as possible.

See Chapter 7 in the *Planning Guide Guide* for complete information on customizing and setting up notification of cluster events.

See the Sample Custom Scripts section in Chapter 1: Troubleshooting HACMP Clusters in the *Troubleshooting Guide* for tips on writing scripts to make **cron** jobs and print queues highly available. There is also a plug-in for print queues in the **usr/es/sbin/cluster/samples** directory.

Customizing Application Server Scripts

See Appendix B: Applications and HACMP in the *Planning Guide* for tips on handling applications.

Some key things to keep in mind:

- Define an HACMP application server for each node that supports applications requiring recovery.
- Applications must be started up and shut down in an orderly fashion. Some situations exist where the timing and control of starting and stopping applications needs to be handled based on pre/post event process. You may need to take into account the order in which applications assigned to the same node are started. Optionally, you can also include applications in different resource groups and establish dependencies between resource groups. For more information, see Adding Resources and Attributes to Resource Groups Using the Extended Path in Chapter 5: Configuring HACMP Resource Groups (Extended).
- Check for dependencies between nodes. For example, a process on node1 may *not* start until a process that runs on node2 is up. Include a check for remote node/application availability before issuing the local startup command.
- You may need to perform some checks to make sure the application is *not* running and to clean up logs or roll back files before starting the application process.

See Sample Custom Scripts section in Chapter 1: Troubleshooting HACMP Clusters in the *Troubleshooting Guide* for tips on writing scripts to make **cron** jobs and print queues highly available. There is also a plug-in for print queues in the **usr/es/sbin/cluster/samples** directory.

Application Monitoring

You can monitor a set of applications that you define through the SMIT interface.

In HACMP 5.2 and up you can configure multiple application monitors and associate them with one or more application servers. By supporting multiple monitors per application, HACMP can support more complex configurations. For example, you can configure one monitor for each instance of an Oracle parallel server in use. Or, you can configure a custom monitor to check the health of the database, and a process termination monitor to instantly detect termination of the database process.

You assign each monitor a unique name in SMIT.

Prior to HACMP 5.2, each application that is kept highly available could have only one of the two types of monitors configured for it. *Process application monitoring* detected the death of one or more processes using RSCT Resource Monitoring and Control (RMC). *Custom application monitoring* checked the health of an application at user-specified polling intervals.

For example, you could supply a script to HACMP that sends a request to a database to check that it is functioning. A non-zero exit from the customized script indicated a failure of the monitored application, and HACMP responded by trying to recover the resource group that contains the application. However, you could *not* use two monitors for one application.

For instructions, see Configuring Multiple Application Monitors in Chapter 4: Configuring HACMP Cluster Topology and Resources (Extended).

With each monitor configured, when a problem is detected, HACMP attempts to restart the application, and continues up to a specified retry count. You select one of the following responses for HACMP to take when an application cannot be restarted within the retry count:

- The **fallover** option causes the resource group containing the application to fall over to the node with the next-highest priority according to the resource policy.
- The **notify** option causes HACMP to generate a **server_down** event, to inform the cluster of the failure.

You can customize the restart process through the Notify Method, Cleanup Method, and Restart Method SMIT fields, and by adding pre- and post-event scripts to any of the failure action or restart events you choose.

Note: If the System Resource Controller (SRC) is configured to restart the application, this can interfere with actions taken by application monitoring. Disable the SRC restart for the application (application start and stop scripts should *not* use the SRC unless the application is *not* restartable). For the case of a custom monitor, the script is responsible for the correct operation. The action taken by application monitoring is supported based on the script return.

Note: If a monitored application is under control of the system resource controller, check to be certain that `action:multi` are **-O** and **-Q**. The **-O** Specifies that the subsystem is *not* restarted if it stops abnormally. The **-Q** Specifies that multiple instances of the subsystem are *not* allowed to run at the same time. These values can be checked using the following command:

```
lssrc -Ss <Subsystem> | cut -d : -f 10,11
```

If the values are *not* **-O** and **-Q** then they must be changed using the **chssys** command.

Measuring Application Availability

You can use the Application Availability Analysis Tool to measure the amount of time that any of your applications (with defined application server) is available. The HACMP software collects, time stamps, and logs the following information:

- An application starts, stops, or fails
- A node fails or is shut down, or comes up

- A resource group is taken offline or moved
- Application monitoring is suspended or resumed.

Using SMIT, you can select a time period and the tool will display uptime and downtime statistics for a given application during that period. The tool displays:

- Percentage of uptime
- Amount of uptime
- Longest period of uptime
- Percentage of downtime
- Amount of downtime
- Longest period of downtime

Note: All nodes must be available when you run the tool to display these statistics. Clocks on all nodes must be synchronized in order to get accurate readings.

See Chapter 4: Configuring HACMP Cluster Topology and Resources (Extended) and Chapter 10: Monitoring an HACMP Cluster, for complete information on application monitoring and configuring and using this tool.

Network Configuration and Nameserving

Setting up and maintaining clear communication paths for the Cluster Manager is a key element for efficient cluster operation.

Setting up Serial Networks or Other Heartbeat Path

It is crucial to have at least one serial network configured for the cluster. Without a serial network, you run the risk of a partitioned cluster if TCP/IP networks fail, since the nodes will be unable to maintain heartbeat communication. You can also use disk heartbeats or heartbeats over IP aliases to maintain cluster communications.

Integrating HACMP with Network Services

HACMP requires IP address to name resolution. The three most commonly used methods include:

- Domain Name Service
- Network Information Service
- Flat file name resolution **(/etc/hosts)**.

By default, a name request will look first for the DNS (**/etc/resolv.conf**), second for NIS, and last for **/etc/hosts** to resolve the name. Since DNS and NIS both require certain hosts as designated servers, it is necessary to maintain the **/etc/hosts** file in case the DNS or NIS name server is unavailable, and to identify hosts that are *not* known to the name server. It is required to have all HACMP IP labels in all cluster nodes' **/etc/hosts** tables.

To ensure the most rapid name resolution of cluster nodes, change the default order for name serving so that **/etc/hosts** is used first (at least for cluster nodes).

To do this, edit the **/etc/netsvc.conf** file so that this line appears as follows:

```
hosts=local,nis,bind
```

Putting the local option first tells the system to use **/etc/hosts** first, then NIS.

You can also change the order for name resolution by changing the environment variable NSORDER as follows:

```
NSORDER=local,bind,nis
```

Note: By default, during the process of IP address swapping, to ensure that the external name service does *not* cause AIX 5L to map the service IP address to the wrong network interface, HACMP automatically disables NIS or DNS by temporarily setting the AIX 5L environment variable `NSORDER=local` within the event scripts.

Note: If you are using NIS, have the NIS master server outside the cluster, and have the cluster nodes run as NIS slave servers. At a minimum, every HACMP node must be able to access NIS master or slave servers on a local subnet, and *not* via a router.

See the *Planning Guide* and the *Installation Guide* for information on editing the **/etc/hosts** file, and also for notes on NIS and **cron** considerations.

WARNING: You cannot use DHCP to allocate IP addresses to HACMP cluster nodes. Clients may use this method, but cluster nodes cannot.

Tuning Networks for Best Performance

HACMP provides easy control over several tuning parameters that affect the cluster's performance. Setting these tuning parameters correctly to ensure throughput and adjusting the HACMP failure detection rate can help avoid "failures" caused by heavy network traffic.

Cluster nodes sometimes experience extreme performance problems, caused by large I/O transfers, excessive error logging, or lack of memory. When this happens, the HACMP daemons can be starved for CPU time. Processes running at a priority higher than the RSCT or Cluster Manager subsystems can also cause this problem.

The deadman switch is an AIX 5L kernel extension that halts a node when the Cluster Manager does *not* run for a certain amount of time, usually due to one of the problems noted above.

See Chapter 9: Configuring AIX 5L for HACMP in the *Installation Guide* for information on setting tuning parameters correctly to avoid some of the performance problems noted above.

If you are running a cluster on an SP, also consult your SP manual set for instructions on tuning SP switch networks.

Planning Disks and Volume Groups

Planning the disk layout is crucial for the protection of your critical data in an HACMP cluster. Follow the guidelines carefully, and keep in mind these issues:

- All operating system files should reside in the root volume group (**rootvg**) and all user data should reside outside that group. This makes updating or reinstalling the operating system and backing up data more manageable.

- A node whose resources are *not* designed to be taken over should *not* own critical volume groups.
- When using copies, each physical volume using a mirror copy should get its power from a UPS system.
- Volume groups that contain at least three physical volumes provide the maximum availability when implementing mirroring (one mirrored copy for each physical volume).
- **auto-varyon** must be set to **false**. HACMP will be managing the disks and varying them on and off as needed to handle cluster events.

Quorum Issues

Setting up quorum correctly when laying out a volume group is very important. Quorum *must* be enabled on concurrent volume groups. With quorum enabled, a two-disk non-concurrent volume group puts you at risk for losing quorum and data access. The failure of a single adapter or cable would cause half the disks to be inaccessible. HACMP provides some protections to avoid the failure, but planning is still important.

Either build three-disk volume groups or disable quorum on non-concurrent volume groups. You can also use the **forced varyon** option to work around quorum issues.

See the detailed section about Quorum in Chapter 5: Planning Shared LVM Components in the *Planning Guide.*

HACMP selectively provides recovery for resource groups that are affected by failures of individual resources. HACMP automatically reacts to a “loss of quorum” LVM_SA_QUORCLOSE error associated with a volume group going offline on a cluster node. If quorum is lost for a volume group that belongs to a resource group on a cluster node, the system checks whether the LVM_SA_QUORCLOSE error appeared in the node’s AIX 5L error log file and informs the Cluster Manager to selectively move the affected resource group.

Note: When the AIX 5L error log buffer is full, new entries are discarded until space becomes available in the buffer and adds an error log entry to inform you of this problem. For information about increasing the size of the error log device driver internal buffer, see the AIX 5L documentation in section Accessing Publications in About This Guide

Note: HACMP launches selective fallover and moves the affected resource group *only* in the case of the LVM_SA_QUORCLOSE error. Be aware that this error occurs if you use mirrored volume groups with quorum enabled. However, in many cases, different types of “volume group failure” errors could occur. HACMP does *not* react to any other type of volume group errors automatically. In these cases, you still need to configure customized error notification methods, or use AIX 5L Automatic Error Notification methods to react to volume group failures.

For more information on Selective Fallover triggered by loss of quorum for a volume group, see the section Selective Fallover Caused by a Volume Group Loss in Appendix B: Resource Group Behavior during Cluster Events.

Planning Hardware Maintenance

Good maintenance practice in general dictates that you:

- Check cluster power supplies periodically
- Check the **errlog** and/or any other logs where you have redirected information of interest and attend to all notifications in a timely manner
- Be prepared to replace any failed or outdated cluster hardware.

If possible, you should have replacement parts readily available. If the cluster has no single points of failure, it will continue to function even though a part has failed. However, now a single point of failure may exist. If you have set up notification for hardware errors, you have an early warning system in place.

This guide contains procedures detailing how to replace the following cluster components while keeping the cluster running:

- Network
- Network interface card
- Disk
- Node.

See Hardware Maintenance for more information.

Planning Software Maintenance

Planning for software maintenance includes:

- Customizing notification of software problems
- Periodically checking and cleaning up log files
- Taking cluster snapshots when making any change to the cluster configuration
- Preparing for upgrading AIX 5L, applications, and HACMP for AIX 5L.

See Preventive Maintenance for more details.

Runtime Maintenance

Once you have configured the cluster and brought it online, it is very important to do maintenance tasks in as non-disruptive a way as possible. Maintaining an HACMP cluster requires attention to some issues that have different ramifications in the cluster environment compared to maintaining a single system.

This section discusses the following issues:

- Tasks that require stopping the cluster
- Warnings about the cascading effects caused by making certain types of changes to a stable, running cluster

Tasks that Require Stopping the Cluster

HACMP allows you to do many tasks without stopping the cluster; you can do many tasks dynamically using the DARE and C-SPOC utilities. However, in order to do the following tasks, you must stop the cluster:

- Change the name of a cluster component: network module, cluster node, or network interface. Once you configure the cluster, you should *not* need to change these names.
- Maintain RSCT.
- Change automatic error notification.
- Change SSA fence registers.
- Convert a service IP label from IPAT via IP Replacement to IPAT via IP Aliases.

Changing the Cluster Configuration—Cascading Effects on Cluster Behavior

Installing HACMP makes changes to several AIX 5L files (see Chapter 1: Administering an HACMP Cluster). All the components of the cluster are under HACMP control once you configure, synchronize, and run the cluster software. Using AIX 5L to change any cluster component, instead of using the HACMP menus and synchronizing the topology and/or the cluster resources, will interfere with the proper behavior of the HACMP cluster software and thus affect critical cluster services.

This section contains warnings about actions that will endanger the proper behavior of an HACMP cluster. It also includes some reminders about proper maintenance procedures.

Stopping and Starting Cluster Services

Do *not* directly start or stop daemons or services that are running under the control of HACMP. Any such action will affect cluster communication and behavior. You can choose to run certain daemons (Clinfo) but others are required to run under HACMP control.

Most important, never use the **kill – 9** command to stop the Cluster Manager or any RSCT daemons. This causes an abnormal exit. SRC will run the **clexit.rc** script and halt the system immediately. This causes the other nodes to initiate a fallover.

TCP/IP services are required for cluster communication. Do *not* stop this service on a cluster node. If you need to stop HACMP or TCP/IP to maintain a node, use the proper procedure to move the node's resources to another cluster node, then stop cluster services on this node. Follow the instructions in Chapter 15: Managing Resource Groups in a Cluster to make changes to cluster topology or resources.

Node, Network, and Network Interface Issues

The HACMP configuration of the nodes and IP addresses is crucial to the communication system of the cluster. Any change in the definitions of these elements must be updated in the cluster configuration and resynchronized.

Do *not* change the configuration of a cluster node, network, or network interface using AIX 5L SMIT menus or commands, individually on a cluster node, outside of HACMP. See Chapter 15: Managing Resource Groups in a Cluster, for instructions on changing the configuration dynamically following the proper HACMP cluster procedures.

Do *not* start or stop daemons or services that are running under the control of HACMP. This action will affect cluster communication and behavior.

Be sure to follow proper procedures for the following types of changes:

- Changing the IP label/address of any network interface defined to HACMP. Changes to IP addresses must be updated in the HACMP cluster definition and the cluster must then be resynchronized. Any change to network interface attributes normally requires stopping cluster services, making the change, and restarting cluster services.

  Note that in some circumstances you can use the HACMP facility to swap a network service IP address dynamically, to another active network interface on the same node and network, without shutting down cluster services on the node. See Swapping IP Addresses between Communication Interfaces Dynamically in Chapter 13: Managing the Cluster Topology for more information.
- Changing netmasks of network interfaces. Service and other network interfaces on the same network must have the same netmask on all cluster nodes. Changes made outside the cluster definition will affect the ability of the Cluster Manager to send heartbeat messages across the network.

  It is important to configure the correct interface name for network interfaces. see the relevant section in Chapter 15: Managing Resource Groups in a Cluster.
- Enabling an alternate Hardware Address for a service network interface using AIX 5L SMIT.
- Taking down network interface cards. Do *not* take down all cards on the same network if a local network failure event is set up to stop cluster services and move resource groups to another node. If the cluster is customized to stop cluster services and move resource groups to another node when all communications on a specific network fail, and you take down all network interfaces, this will force the resource groups to move to another node whether you intend it or *not*.
- Taking down network interfaces. Do *not* bring all network interfaces down on the same network if there is only one network and no point-to-point network is defined. Doing this will cause system contention between cluster nodes and fallover attempts made by each node. A Group Services domain merge message is issued when a node has been out of communication with the cluster and then attempts to reestablish communication. The cluster will remain unstable until you fix the problem.

Making Changes to Network Interfaces

In some circumstances, you can use the HACMP facility to swap a network service IP address dynamically, to an active standby network interface on the same node and network, without shutting down cluster services on the node. See Swapping IP Addresses between Communication Interfaces Dynamically in Chapter 13: Managing the Cluster Topology for more information.

Typically, stop the cluster to make any change to network interfaces. If you must change the IP address of an network interface, or if you change the IP label/address, make sure to make the changes to both DNS or NIS *and* the **/etc/hosts** file. If DNS or NIS and **/etc/hosts** are *not* updated, you will be unable to synchronize the cluster nodes or do any DARE operations. If DNS or NIS services are interrupted, the **/etc/hosts** file is used for name resolution. You must also redo **cl_setup kerberos** if you are using Kerberos security.

Handling Network Load/Error rates

Dropped packets due to network loads may cause false fallovers. Also, high throughput may cause the deadman switch to time-out. If either of these conditions occurs, check the AIX 5L network options and the Failure Detection Rate you have set for the cluster. These parameters are contained in the **Advanced Performance Tuning Parameters** panel in SMIT.

See Changing the Configuration of a Network Module in Chapter 13: Managing the Cluster Topology, for information on tuning the Failure Detection Rate. RSCT logging can also help with the tuning of networks.

Maintaining and Reconfiguring Networks

Moving Ethernet ports on a running cluster results in network interface swap or node failure. Even a brief outage results in a cluster event.

Shared Disk, Volume Group, and Filesystem Issues

Do *not* change the configuration of an HACMP shared volume group or filesystem using AIX 5L, outside of HACMP. Any such action will affect cluster behavior. The Cluster Manager and the cluster event scripts assume the shared volume groups and filesystems are under HACMP control. If you change the environment, the event scripts will *not* be able to complete properly and you will get unexpected results.

Disk Issues

Disks should always be mirrored (or use a disk array), to protect against loss of data. Once they are defined and configured within the HACMP cluster, you should always use the HACMP C-SPOC utility (`smit cl_admin`) to add or remove disks from a volume group with the cluster running. The cluster needs to be made aware of disks being added to or removed from a shared volume group. If you add or remove disks using the conventional method, the cluster will *not* be aware that these changes have occurred.

Volume Group and Filesystems Issues

Use the C-SPOC utility (`smit cl_admin`) for common maintenance tasks like creating, extending, changing, or removing a shared filesystem. See Chapter 11: Managing Shared LVM Components.

See Chapter 5: Planning Shared LVM Components in the *Planning Guide* for information on using NFS and HACMP.

When configuring volume groups and filesystems:

- Do *not* set filesystems to automount; HACMP handles the mounts at startup and during cluster events.
- Do *not* set volume groups to autovaryon; HACMP executes the varying on and off as needed.
- If you are testing something when the cluster is *not* running and you varyon a volume group or mount a filesystem, remember to unmount the filesystem and vary off the volume group before you start HACMP.
- Do *not* have any processes running that would point to a shared filesystem when cluster services are stopped with resource groups brought offline on the node that currently owns that filesystem. If cluster services are stopped with resource groups brought offline and the application stop script fails to terminate the processes that are using the filesystem, that filesystem will be unable to unmount and the fallover will *not* occur. The cluster will go into a **config_too_long** condition.

One of the more common reasons for a filesystem to fail being unmounted when cluster services are stopped with resource groups brought offline is because the filesystem is busy. In order to unmount a filesystem successfully, no processes or users can be accessing it at the time. If a user or process is holding it, the filesystem will be "busy" and will *not* unmount. The same issue may result if a file has been deleted but is still open.

This is easy to overlook when you write application stop scripts. The script to stop an application should also include a check to make sure that the shared filesystems are *not* in use. You can do this by using the **fuser** command. The script should use the **fuser** command to see what processes or users are accessing the filesystems in question. These processes can then be killed. This will free the filesystem so it can be unmounted.

Refer to the AIX 5L man pages for complete information on this command.

Expanding Filesystems

Use C-SPOC to increase the size of a filesystem:

1. Enter `smit cl_admin`
2. Go to **System Management (C-SPOC) > HACMP Logical Volume Management > Shared Filesystems > JFS or Enhanced JFS** (depending on the case) and press Enter.
3. Select the option to change a cluster filesystem.
4. Select the filesystem to change.
5. Enter the new size for the filesystem.
6. Return to the **Logical Volume Management** panel and synchronize the new definition to all cluster nodes via **Synchronize a Shared Volume Group Definition**.

General Filesystems Issues

The following are some more general filesystems concerns:

- Full filesystems in the root volume group may cause cluster events to fail. You should monitor this volume group and clean it up periodically. You can set up a **cron** job to monitor filesystem size to help avoid filling a critical filesystem (for example, the **hacmp.out** file can get quite large).
- Shared filesystems must have the *mount* option set to false, so that HACMP can mount and unmount them as needed to handle cluster events.
- Be aware of the way NFS filesystems are handled. See Using NFS with HACMP in Chapter 14: Managing the Cluster Resources.

For information on using GPFS, see Appendix C in the *Installation Guide*.

Application Issues

Appendix B: Applications and HACMP in the *Planning Guide* gives many pointers on planning and maintaining applications in the HACMP environment. Some key points to remember:

- Application maintenance will require downtime for resource groups if binaries reside on a shared disk.
- Upgrades should be tested prior to implementation to anticipate effects on the production cluster.
- Changes to application start and stop procedures should be thoroughly tested prior to going into production.

- Do *not* have shared applications already running when you start the cluster. A second attempt at starting already running applications may cause a problem.
- Do *not* manually execute the application stop script for any reason on a running cluster without starting the application back up again. Problems may occur if an attempt is made to stop the application that is already down. This could potentially cause a fallover attempt to be unsuccessful.

Hardware Maintenance

Hardware failures must be dealt with promptly, as they may create single points of failure in the cluster. If you have carefully set up error notification and event customization as recommended, you receive quick notification via email of any problems. You should also periodically do error log analysis. See Viewing HACMP Cluster Log Files section in Chapter 2: Using Cluster Log Files in the *Troubleshooting Guide* for details on error log analysis.

Some issues to be aware of in a high availability environment include:

- Shared disks connect to both systems, thus open loops and failed disks can result in fragmented SSA loops and the loss of access to one mirror set.
- Set up mirroring so that the mirrored disk copy is accessible by a different controller. This prevents loss of data access when a disk controller fails. When a disk controller fails, the mirror disk is accessible through the other controller.

System ID Licensing Issues

The Concurrent Resource Manager is licensed to the system ID of a cluster node. Many of the **clvm** or concurrent access commands validate the ID against the license file. A mismatch will cause the command to fail, with an error message indicating the lack of a license.

Restoring a system image from a **mksysb** tape created on a different node or replacing the planar board on a node will cause this problem. In such cases, you must recreate the license file by removing and reinstalling the **cluster.clvm** component of the current release from the original installation images.

Replacing Topology Hardware

Nodes, networks, and network interfaces and devices comprise the topology hardware. Changes to the cluster topology often involves downtime on one or more nodes if changes to cabling or adding/removing network interfaces is involved. In most situations, you can use the DARE utilities to add a topology resource without downtime.

The following sections indicate the conditions under which you can use DARE and the conditions under which you must plan cluster downtime.

Note: No automatic corrective actions take place during a DARE.

Replacing Nodes

Using the DARE utility, you can add or remove a node while the cluster is running

Replacing a Node or Node Component

If you are replacing a cluster node keep this list in mind:

- The new node must typically have the same amount of RAM (or more) as the original cluster node.
- The new node must typically be same type of system if your applications are optimized for a particular processor.
- The new node's slot capacity typically must be the same or better than the old node.
- NIC physical placement is important – use the same slots as originally assigned.
- If you have a concurrent environment, you must reinstall the CRM (Concurrent Resource Manager) software. This is also a consideration if you are cloning nodes.
- Get the new license key from the application vendor for the new CPU ID if necessary.

If you are replacing a component of the node:

- Be aware of CPU ID issues
- For SCSI adapter replacement – reset external bus SCSI ID to original SCSI ID
- For NIC replacement – use the same slots as originally assigned.

Procedure for Adding or Removing a Node

Also see Changing the Configuration of Cluster Nodes in Chapter 13: Managing the Cluster Topology for more complete information.

The basic procedure for adding or removing a node:

1. Install AIX 5L, HACMP and LPPs on new node and apply PTFs to match the levels of the previous node.
2. Connect networks and SSA cabling and test.
3. Configure TCP/IP.
4. Import volume group definitions.
5. Connect serial network and test.
6. Change the Configuration Database configuration on one of the existing nodes.
7. Synchronize and verify from the node where you made the changes.

Replacing Networks and Network Interfaces

You can only protect your applications from downtime due to a network failure if you configure more than one IP network. You should also have a serial network. (or use heartbeat over disk or heartbeat over IP aliases). If no backup network is configured, the cluster will be inaccessible to all but directly connected clients.

Note: It is important to configure the correct interface name for network interfaces. See Chapter 15: Managing Resource Groups in a Cluster.

You can replace network cabling without taking HACMP off line. You can also replace hubs, routers, and bridges while HACMP is running. Be sure to use the correct IP addresses when reconfiguring a router.

You can use the DARE **swap_adapter** function to swap the IP address on the same node and network. Then you can service the failed network interface card without stopping the node.

Procedure for Replacing a LAN adapter

If the hardware supports hot-pluggable network interfaces, no cluster downtime is required for this procedure.

If you cannot use the **swap_adapter** function, use this procedure:

1. Move resource groups to another node using the Resource Group Management utility.
2. Use the hotplug mechanism to replace the card.
3. Assign IP addresses and netmasks for interfaces if they were undefined (see Configuring Cluster Topology (Extended) in Chapter 4: Configuring HACMP Cluster Topology and Resources (Extended).
4. Test IP communications.

Handling Disk Failures

Handling shared disk failures differs depending on the type of disk and whether it is a concurrent access configuration or *not*.

- SCSI non-RAID—You will have to shut down the nodes sharing the disks.
- SCSI RAID—There may be downtime, depending on the capabilities of the array.
- SSA non-RAID—Requires manual intervention – you can replace disks with no system downtime.
- SSA RAID—No downtime necessary.

See Restarting the Concurrent Access Daemon (clvmd) in Chapter 12: Managing Shared LVM Components in a Concurrent Access Environment for information on that procedure.

Replacing a Failed SSA non-RAID Disk

Use C-SPOC to replace a mirrored disk drive. See Maintaining Physical Volumes in Chapter 11: Managing Shared LVM Components.

Preventive Maintenance

If you have a complex and/or very critical cluster, it is highly recommended that you maintain a test cluster as well as your production cluster. Thus before you make any major change to the production cluster, you can test the procedure on the test cluster.

HACMP also supplies event emulation utilities to aid in testing.

Cluster Snapshots

Periodically take snapshots of the cluster in case you need to reapply a configuration. You should take a snapshot any time you change the configuration. Keep a copy of the snapshot on another system, off the cluster, as protection against loss of the cluster configuration. You can use the snapshot to rebuild the cluster quickly in case of an emergency. See Chapter 18: Saving and Restoring Cluster Configurations in this Guide for complete information on cluster snapshots. You might want to consider setting up a **cron** job to do this on a regular basis.

Backups

HACMP does *not* provide tools for backing up the system. You should plan for periodic backups just as you do for a single system. You should do backups of **rootvg** and shared volume groups.

Backups of shared volume groups should be done frequently.

Some applications have their own online backup methods.

You can use any of the following:

- **mksysb** backups
- Online backups (**sysback, splitlvcopy**)

Using mksysb

You should do a **mksysb** on each node prior to and following any changes to the node environment. Such changes include:

- Applying PTFs
- Upgrading AIX 5L or HACMP software
- Adding new applications
- Adding new device drivers
- Changing TCP/IP configuration
- Changing cluster topology or resources
- Changing LVM components of **rootvg** (paging space, filesystem sizes)
- Changing AIX 5L parameters (including the tuning parameters: I/O pacing, **syncd**)

Using splitlvcopy

You can use the **splitlvcopy** method on raw logical volumes and filesystems to do a backup while the application is still running. This method is only possible for LVM mirrored logical volumes.

By taking advantage of the LVM's mirroring capability, you can stop the application briefly to split off a copy of the data using the AIX 5L **splitlvcopy** command. Stopping the application gives the application its checkpoint. Then restart the application so it continues processing while you do a backup of the copy.

You can do the backup using **tar**, **cpio**, or any other AIX 5L backup command that operates on a logical volume or a filesystem. Using **cron**, you can automate this type of backup.

Using cron

Use the AIX 5L **cron** utility to automate scheduled maintenance and to monitor the system.

Using cron to Automate Maintenance of Log Files

Use this utility to automate some of the administrative functions that need to be done on a regular basis. Some of the HACMP log files need **cron** jobs to ensure that they do *not* use up too much space.

Use **crontab –e** to edit **/var/spool/cron/crontabs/root**.

Cron will recognize the change without need for rebooting.

You might establish a policy for each log, depending how long you want to keep the log, and what size you will allow it to grow. **hacmp.out** is already set to expire after it cycles >7 times.

The RSCT logs are stored in the **/var/ha/log** directory. These logs are trimmed regularly. If you want to save information for a longer period of time you can either redirect the logging to a different directory, or change the maximum size file parameter (using SMIT). See Viewing HACMP Cluster Log Files section in Chapter 2: Using Cluster Log Files in the *Troubleshooting Guide*.

Using cron to Set up An Early Warning System

Use **cron** to set up jobs to proactively check out the system:

- Run a custom verification daily and send a report to the system administrator.
- Check for full filesystems (and take action if necessary).
- Check that certain processes are running.
- Run event emulation and send a report to the system administrator.

Do Regular Testing

Regularly schedule a testing window where a failure is conducted in a controlled environment. That way you can evaluate a fallover before anything happens in your production cluster. It should include fallovers of all nodes and full verification of tested protected applications. This is strongly encouraged if you are changing or evolving your cluster environment.

Upgrading Software (AIX 5L and HACMP)

When upgrading the AIX 5L or HACMP software:

- Take a cluster snapshot and save it in a directory outside the cluster.
- Back up the operating system and data before performing any upgrade. Prepare a backout plan in case you encounter problems with the upgrade.
- Whenever possible, plan and do an initial run through on a test cluster.
- Use disk update if possible.
- Follow this same general rule for fixes to the application; follow specific instructions for the application.

AIX 5L fixes need to be applied according to the HACMP operations guide:

- Apply APARs to standby node.
- Fallover (stopping cluster services with the Move Resource Groups option) to standby machine.
- Apply APARs.

See the *Installation Guide* for installation and migration procedures.

Appendix B: Resource Group Behavior during Cluster Events

This appendix describes how HACMP manages resource groups. It contains information on the following:

- Resource Group Event Handling and Recovery
- General Resource Group Event Processing Logic
- Selective Fallover for Handling Resource Groups
- Handling of Resource Group Acquisition Failures
- Recovering Resource Groups when Nodes Join the Cluster
- Handling of Resource Groups Configured with IPAT via IP Aliases
- Examples of Location Dependency and Resource Group Behavior.

The information in this appendix may be especially useful for experienced HACMP users who are familiar with previous resource group handling by HACMP but may *not* be aware of changes made in recent releases.

It is assumed that the reader is familiar with basic resource group fallover policies. These concepts are covered in the *Concepts Guide* and the *Planning Guide*.

Overview

Once you have planned and defined the cluster topology and resources, HACMP monitors the working cluster and takes actions to recover when failures are detected. HACMP monitors resources and launches events to handle resource groups. You *do not* need to specify these actions in the cluster; they are initiated automatically.

HACMP manages resource groups by:

- Moving only the resource groups that are affected by a failure of an individual resource to another node in the cluster.
- Taking recovery actions when it fails to acquire resource groups on a node. This can happen when HACMP attempts to move a resource group to another node in the cluster, but fails to acquire it on that node. You can disable automatic recovery, if needed.

The sections below provide an overview of resource group events and describe when HACMP moves resource groups in the cluster, how the resource groups are placed on the nodes, and how to identify the causes of the underlying cluster events.

Resource Group Event Handling and Recovery

HACMP tracks the state of all cluster resources and manages the recovery according to the available backup resources. When multiple backup resources are available, HACMP can be configured to dynamically select the backup resources to use based on the current performance statistics (using a dynamic node priority policy). Event logging includes a detailed summary for each high-level event to help you understand exactly what actions were taken for each resource group during the handling of failures.

Events and Resource Groups with Dependencies or Sites

If either parent/child or location dependencies between any resource groups or sites are configured in the cluster, HACMP processes all events related to resource groups in the cluster with the use of **resource_state_change** trigger events that are launched for all resource groups for events where resource groups are affected.

The Cluster Manager then takes into account all configured runtime policies, especially the configuration of dependencies and sites for resource groups, and the current distribution and state of resource groups on all nodes in order to properly handle any acquiring, releasing, bringing online or taking offline of resource groups.

When events handle resource groups with dependencies or sites, a preamble is written to the **hacmp.out** log file listing the plan of sub_events for handling the resource groups.

resource_state_change This trigger event is used for resource group recovery if resource group parent/child or location dependencies or sites are configured in the cluster. This action indicates that the Cluster Manager needs to change the state of one or more resource groups, or there is a change in the state of a resource managed by the Cluster Manager. This event runs on all nodes if one of the following occurs:

- Application monitoring failure
- Selective fallover for loss of volume group
- Local network down
- WAN failure
- Resource Group Acquisition Failure
- Resource Group Recovery on IP Interface Availability
- Expiry of settling timer for a resource group
- Expiry of fallback timer for a resource group

While the event runs, the state of the resource group is changed to TEMP_ERROR or SECONDARY_TEMP_ERROR. This is broadcast to all nodes.

NOTE: This is a place where you can add pre- or post-events for specific resources if needed.

resource_state_change_complete This event runs on all nodes when the **resource_state_change** event has successfully completed (necessary recovery actions including release and acquire events have completed).

Events for Moving Resource Groups

HACMP may move resource groups as a result of recovery actions taken during the processing of events such as **node_down** and especially for **resource_state_change**:

rg_move	This event moves a specified resource group from one node to another.
rg_move_complete	This action indicates that the **rg_move** event has successfully completed.

Resource Group Subevents and States

Handling of individual resources during the processing of an event may include the following actions or resource group states. For example, when a filesystem is in the process of being unmounted and mounted it is taken offline and then released by one node. Then if there is an available backup node the filesystem will be acquired and brought online.

RELEASING	A resource group is being released either to be brought offline, or to be acquired on another node.
ACQUIRING	A resource group is being acquired on a node.
ONLINE	The resource group is online.
OFFLINE	The resource group is offline.
ERROR	The resource group is in an error state.
TEMPORARY ERROR	The resource group is in a temporary error state. It occurs, for instance, due to a local network failure or an application failure. This state informs the Cluster Manager to initiate an **rg_move** event for this resource group. Resource groups should *not* be in this state when the cluster is stable. See the section Resource Group Recovery when the Network or Interface is Up in this appendix.
UNKNOWN	The state of the resource group is unknown.
ONLINE SECONDARY	The resource group is online at the secondary site.
ONLINE PEER	The resource group is online on a peer node.
ACQUIRING SECONDARY	A resource group is being acquired on a node at the secondary site.
RELEASING SECONDARY	A resource group is being released on a node at the secondary site.
ERROR SECONDARY	The resource group is in error state at the secondary site.

TEMPORARY ERROR SECONDARY	The resource group is in a temporary error state at the secondary site.
UNMANAGED	You have stopped the cluster services without stopping the running applications. In this case: • HACMP is *not* managing the resources. • The previous state of the group was ONLINE. • The application and other resources may continue to be *running* on the node.

Note: See the *Planning Guide*, Chapter 6, Planning Resource Groups, for more information on resource group behavior in clusters with sites.

After the completion of an event, HACMP is aware of the state of resources and resource groups involved in the event. HACMP then analyzes the resource group information that it maintains internally and determines whether recovery events need to be queued for any of the resource groups. HACMP also uses status of individual resources in resource groups to print out a comprehensive event summary to the **hacmp.out** log file.

For each resource group, HACMP keeps track of the nodes on which the resource group has tried to come online and failed. This information is updated when recovery events are processed. HACMP resets the nodelist for a resource group as soon as the resource group moves to the online or error states.

When a resource group is in the process of being moved, application monitoring is suspended and resumed appropriately. The Application Monitor sees that the application is in *recovery* state while the event is being processed.

resume_appmon	This action is used by the Application Monitor to resume monitoring of an application.
suspend_appmon	This action is used by the Application Monitor to suspend monitoring of an application.

Cluster Event Processing

The resource group handling features add steps to the overall processing for an event:

1. The Cluster Manager communicates with RSCT Group Services to obtain information about topology events, and consolidates information about events related to resource groups.
2. The Cluster Manager performs event rollup and determines the actual cluster event to run.
3. The Group Services protocol is run to get all cluster nodes to agree on the event (voting).
4. The Cluster Manager starts up event scripts on the cluster nodes.
5. Event scripts get information about resource groups to process for the event:
   - Get information from the HACMP Configuration Database and the Cluster Manager and determine resource groups to process for the event.

- Get information about the nodes already tried and exclude these nodes from the default nodelist.
- Exclude nodes with insufficient network interfaces (for resource groups that require network interfaces).

6. Check the node priority policy to prioritize the list of target nodes for a resource group.
7. Event scripts process resource groups (bring them online/offline, etc.).
8. Resource Group Manager internally marks resource groups as recoverable if failures are encountered during the “acquire” phase.
9. Event scripts complete.

Note: For more information on steps 5-9, and for information on which attributes and policies take precedence when the Cluster Manager determines which resource group to process first, see the section General Resource Group Event Processing Logic.

10. The Cluster Manager gets return code from scripts.
11. If the return code is **0**, the event has completed (it may or may *not* have been successful); otherwise, return **event_error** (user intervention may be required to get the cluster back to a stable state).
12. The Cluster Manager notes the resource groups affected by a local network failure event and marks affected resource groups as recoverable.
13. The Cluster Manager notes the resource groups affected by a local network failure event (13) or by an acquiring error (8) and enqueues recovery events for each resource group in the recoverable state.
14. End of event.

General Resource Group Event Processing Logic

You can specify various policies for resource groups in the cluster that affect the order in which resource group events take place. Such policies may include:

- Specifying customized serial resource group processing. (This may be deprecated in a future release).
- Requesting HACMP to move a resource group to a particular destination node or state. For more information, see the sections on migrating resource groups in Chapter 15: Managing Resource Groups in a Cluster.
- Setting a dynamic node priority.
- Specifying parent/child or location dependencies between resource groups
- Customizing resource recovery for service IP labels and volume group resources (specifying fallover or notify)
- Customizing resource group cross-site recovery (specifying **fallover to other site** or **notify**).

This section presents a high-level view of what actions take precedence in the resource group event processing process. The Cluster Manager examines the following variables to determine the order in which to handle events for resource groups. If all variables are equally true for two or more groups, groups are sorted according to the processing order specified on the nodes in the cluster.

1. Resource group states (online, offline, error, unmanaged) determine which policies will be considered and, therefore, which resource groups will be taken for processing first. For instance, if the resource group is offline, the dynamic node priority set for this resource group is *not* considered. If the resource group is online, the Cluster Manager does *not* have to move it and does *not* perform look-ahead process to find an appropriate node.

   Also, in this step, resource groups' dependencies are considered to determine which resource groups must be processed before other resource groups can be processed.
2. Look-ahead for resource availability is considered. Nodes with insufficient network interfaces (for resource groups that require network interfaces) are excluded, and this affects the order in which events for resource groups will be handled.
3. The node distribution policy for resource groups is considered.
4. Dynamic node priority is considered.
5. Participating nodelist for the particular resource group is considered.
6. Startup, fallover and fallback settings of resource groups are considered: These settings include any previously configured delayed fallback timers and settling times for resource groups.
7. Once the Cluster Manager decides which resource group to process, it considers the resource group dependencies, processing order settings and inter-site management policies. At this step, the Cluster Manager chooses the path for processing.

   Resource groups in clusters with dependencies or sites are processed in phases since more variables must be considered.

Selective Fallover for Handling Resource Groups

Selective fallover is a function of HACMP, which attempts to selectively move only a resource group that has been affected by an individual resource failure to another node in the cluster, rather than moving all resource groups. Selective fallover provides recovery for individual resource groups that are affected by failures of specific resources.

Resources for which Selective Fallover is Used

HACMP utilizes selective fallover in the case of a failure of several types of resources that can belong to a resource group:

- Service IP labels

  For more information, see the section Selective Fallover Caused by Network Interface Failures and the section Selective Fallover Caused by Local Network Failures in this appendix.
- Applications

  For more information, see the section Selective Fallover Caused by Application Failures in this appendix.
- Communication Links.

  For more information, see the section Selective Fallover Caused by a Communication Link Failure in this appendix.
- Volume groups.

  For more information, see the section Selective Fallover Caused by a Volume Group Loss in this appendix.

You can customize the default selective fallover behavior to use a notification instead of a fallover for the following types of resources:

- Service IP labels
- Volume groups. This is the only resource type where customization also affects the secondary instance (if the resource group is replicated).

Selective Fallover and Resource Groups with Sites

If you have configured replicated resource groups, HACMP by default tries to recover the failure of a secondary instance as well as the primary instance. Recovering the secondary instance does *not* affect the primary resource group.

You can change the resource group recovery policy for replicated resource groups to allow (or *not* allow) the Cluster Manager to move the *primary instance* of resource groups to another site in cases where it can use selective fallover to avoid having the resource group go into ERROR state. If the primary instance is moved, then the secondary instances are placed on the opposite site.

For information on customizing resource and cross-site resource group recovery, see Customizing Resource Recovery in Chapter 4: Configuring HACMP Cluster Topology and Resources (Extended).

Look-Ahead for Moving Resource Groups and Choosing Takeover Nodes

HACMP determines the highest priority node using the resource group nodelist, dynamic node priority and persistent migration requests, as well as the availability of backup resources. For example, if a group contains a service IP label, HACMP looks at the status of the available interfaces on the backup nodes. If the resource group is part of a resource group parent/child or location dependency set, HACMP takes this into account also.

HACMP does *not* move a resource group when there are no available backup resources. Instead, it simply takes the group offline from the current node. This result is clearly indicated in the event summary in the **hacmp.out** log file.

Selective Fallover Caused by Network Interface Failures

When a network interface with an HACMP service IP label fails and there are no other network interfaces available on the node on the same HACMP network, the affected applications on that node cannot run. If the service network interface is the last one available on the node, the network interface failure triggers a network failure event.

HACMP distinguishes between two types of network failure, *local* and *global*. A local network failure occurs when a node can no longer communicate over a particular network, but the network is still in use by other nodes. A global network failure occurs when all nodes lose the ability to communicate over a network.

HACMP uses the following formats for local and global network failure events:

Local Network Failure Event	`network_down <node_name> <network_name>`
Global Network Failure Event	`network_down -1 <network_name>`

In the case of a local network failure, you may create a post-event to trigger a **node_down** event. While this has the desired effect of moving the resource group with the failed resource to another node, it has the undesired effect of moving all of the resource groups on the node to other nodes.

Selective fallover uses this infrastructure to better handle network interface failures. You *do not* have to create a post-event to promote a local network failure to a node failure in this case. See the section below for more information on how HACMP handles network interface failures.

You should *not* promote global network failures to **node_down** events as the global network event applies to all nodes and would result in a node down for all nodes.

Actions Taken for Network Interface Failures

HACMP takes the following actions in cases of network interface failures:

- When a network interface with a service IP label fails, and there are no network interfaces available on the same node (therefore, a **swap_adapter** is *not* possible), it moves only the resource group associated with the failed service network interface to another node.
- When a network interface fails and this can result in launching an **rg_move** for the affected resource group, a check for available network interfaces is made. The highest priority node with an available network interface attempts to acquire the resource group.

- HACMP checks that a network interface is available on the node joining the cluster before releasing the resource group. If no network interfaces are available the resource group is *not* released.
- In addition, see the section Resource Group Recovery when the Network or Interface is Up in this chapter for information on when HACMP makes an attempt to bring the resource group that went into an error state back online.

The above actions assume available nodes in the resource group definitions.

The **hacmp.out** file contains messages informing you about cluster activity that results from selective fallover actions.

Selective Fallover Caused by Local Network Failures

When a local network failure event occurs, the Cluster Manager takes selective recovery actions for resource groups containing a service IP label connected to that network. The Cluster Manager tries to move only the resource groups affected by the local network failure event, rather than all resource groups on a particular node.

Note: You *do not* need to create a post-event to promote a local network failure to a node failure in cases of network interface failures.

For example, if you have two resource groups:

```
RG1 - service label on network ether
RG2 - service label on network Token-Ring
```

If network Token-Ring fails, the Cluster Manager will move RG2; it will *not* touch RG1.

Selective Fallover Caused by Application Failures

When an application that is being monitored by Application Monitoring fails, HACMP attempts to move the resource group containing that application to another node. Only the affected resource group is moved.

HACMP can only detect the application failure if an application monitor is configured. There are two types of application monitors:

- Process monitors. You tell HACMP about a specific process to watch. When that process exits, HACMP initiates recovery.
- Custom monitors. You provide your own method, which is called periodically by HACMP to check the health of the application. Custom monitors can be much more sophisticated and robust in their checking of application health.

Selective Fallover Caused by a Communication Link Failure

An X.25 or SNA-over-X.25 communication link that is defined as a resource in a cluster resource group has a list of physical adapters on that node that it can use for reestablishing a connection.

When **clcommlinkd** detects the failure of an X.25 link over which X.25 or SNA is running, the link falls over to another adapter on the same node if possible. If there are no other adapters available on the local node, the selective fallover action moves the resource group containing the link to another node.

An **rg_move** event launched in this case logs the reasons why it was launched in the **hacmp.out** file. The **hacmp.out** file also indicates resource group status.

Selective Fallover Caused by a Volume Group Loss

Selective fallover can also be triggered when HACMP detects a volume group failure on a node containing that resource group. In other words, HACMP automatically reacts to a “loss of quorum” error associated with a volume group going offline on a cluster node.

If a volume group in the resource group has dropped offline due to a loss of quorum error for the volume group on that node, HACMP selectively moves the resource group to another node.

Conditions for Selective Fallover for Volume Group

HACMP uses the selective fallover for volume group loss functionality under the following conditions:

- HACMP monitors all volume groups that are included in the resource group, and all volume groups on which filesystems that are part of the resource group depend.
- HACMP moves only resource groups that contain volume groups with volume groups for which the LVM_SA_QUORCLOSE error has been logged by the error daemon in the AIX 5L **errpt** on that node.

  Note: HACMP does *not* react to any other type of volume group errors automatically. In these cases, you still need to configure customized error notification methods, or use AIX 5L Automatic Error Notification methods to react to volume group failures.

Notes on the Error Notification Method for Volume Group Loss

HACMP uses an Error Notification method to inform the Cluster Manager about the failure of a volume group. When using this error notification method:

- Do *not* modify this error notification method. HACMP issues a warning and takes no action if you attempt to customize this notification method, or to use it to protect against the failure of other types of resources.
- Synchronize the cluster after making changes to the cluster configuration. A notification script used for a volume group failure should correspond to the current configuration of cluster resources, otherwise HACMP issues a warning during verification and takes no action to selectively move the affected resource group.
- Besides the **errnotify** entries created by HACMP for selective fallover, the **errnotify** ODM may also contain other entries related to the same AIX 5L error labels and resources. However, selective fallover provides the most effective recovery mechanism to protect a resource group from the failure of a single resource.
- The notification method that is run in the case of a volume group failure provides the following information in the **hacmp.out** and **clstrmgr.debug** log files:
  - AIX 5L error label and ID
  - The name of the affected resource group
  - The node’s name on which the error occurred.
- You can test the error notification methods generated by the selective fallover facility by emulating an error for each volume group in SMIT. To test error notification:

1. Enter `smit hacmp`
2. In SMIT, select **Problem Determination Tools > HACMP Error Notification > Emulate Error Log Entry**, and press Enter.
3. Select from the picklist the error notification object that was generated by the selective fallover facility for each volume group.

For more information on how the error log emulation works, see Chapter 9: Configuring AIX 5L for HACMP in the *Installation Guide.*

Handling of Resource Group Acquisition Failures

HACMP uses event scripts to move resources around the HACMP cluster. HACMP differentiates certain types of failures in the event script. There are still fatal type errors where an error in the script logic or environment causes the script to fail, but now HACMP traps recoverable errors related to the processing of the resources. This allows HACMP to continue event processing and to try to bring the group online on the next available node.

Attempts by HACMP to start or move a resource group may fail for a variety of reasons, such as busy or unavailable devices, or lack of disk space. HACMP may react to such failures by attempting to move the resource group to another node.

In the event of a resource group acquisition failure on a particular node:

- *Not* all resource group acquisition failures require immediate manual intervention. In some cases, a resource group will be successfully brought online on another node. However, the fact that a resource group acquisition failure occurred indicates a system problem that needs attention.
- The Cluster Manager logs error messages when a node cannot acquire a resource group, and continues processing events so the cluster resources remain available.

  HACMP 5.2 and up *automatically* attempts to activate resource groups in the ERROR state on the node during a **node_up** event. You cannot disable this functionality. If an attempt to recover a resource group in the ERROR state on a joining node is made but the resource group acquisition on the node fails, the non-concurrent resource group falls over to the next node in the nodelist, if one is available. If the concurrent resource group acquisition fails, the resource group remains in the ERROR state.
- HACMP logs reported resource group acquisition failures (failures indicated by a non-zero exit code returned by a command) in **hacmp.out**. The information appears in an event summary that follows each top-level event's details.

  The event summary makes it easier for you to check the **hacmp.out** file for errors. Checking this log becomes more important, since the **config_too_long** console message will *not* be evident in every case where a problem exists.

  However, if you have previously configured a notification method for a **config_too_long** event, HACMP uses this method to notify you that the resource group went offline or remained in an error state.

How HACMP Processes Resource Group Acquisition Failures

If a node fails in an attempt to acquire a resource group during a fallover, HACMP marks the resource group “recoverable” and triggers an **rg_move** event to try to bring up the resource group on some other node. Note that a failure may occur during the acquisition phase of an **rg_move**, and this may cause a queue of **rg_move** events. The software goes through the queue until the resource group is successfully brought online, or until all possible owners have failed to acquire it, in which case the resource group is left in the ERROR state.

Resource Group Recovery when the Network or Interface is Up

When a local network failure occurs, HACMP determines whether any resource groups are affected by the failure. The affected resource groups are those that contain service labels defined on the failed network. In this case, HACMP checks whether such resource groups are online on a node, and attempts to move each affected resource group to another node by initiating an **rg_move** event for it.

The **rg_move** event attempts to bring the resource group online on another node. If HACMP does *not* find available resources on any nodes to bring the resource group online, then the **rg_move** event leaves the resource group in an error state and the resource group goes offline and becomes unavailable.

HACMP tries to bring resource groups in ERROR state online whenever a network interface becomes available. When bringing the affected resource groups online, HACMP does the following:

1. If it finds resource groups that went to an error state due to a resource failure and contain a service IP label of a network interface that became available again, then it moves such resource groups to the **rg_temp_error_state**.
2. Prior to running an **rg_move** for an affected resource group, HACMP determines possible candidate nodes for bringing the resource group online. If it cannot find any candidate nodes, then the resource group remains offline and unavailable.
3. If HACMP finds candidate nodes, it initiates an **rg_move** event to attempt to bring the resource groups online.

Disabling Automatic Recovery of Resource Groups

When a local network failure occurs, you may need to replace the failed resource first, before letting HACMP automatically bring the affected resource group online. For instance, you may need to replace and test the network interface before bringing the affected resource group online.

To avoid automatic recovery of a resource group if the resource group goes into an error state, take these steps:

1. Enter `smit hacmp`
2. In SMIT, select **System Management (C-SPOC) > Resource Group and Application Management > Bring a Resource Group Offline** and press Enter.
3. Specify that this resource group must remain offline on a node.

Remember to bring the resource group back online manually when needed.

Recovering Resource Groups when Nodes Join the Cluster

Prior to HACMP 5.2, when a node joined the cluster, it did *not* make an attempt to acquire any resource groups that had previously gone into an ERROR state on any other node. Such resource groups remained offline and required use of the Resource Group Migration utility, **clRGmove**, to manually bring them back online.

In HACMP 5.2 and up, resource group recovery is improved. An attempt is made to automatically bring online the resource groups that are currently in the ERROR state. This further increases the chances of bringing the applications back online. When a node that is included in the nodelist for the resource group starts up, if the resource group is in the ERROR state on any node in the cluster, this node attempts to acquire the resource group. The node must be included in the nodelist for the resource group.

The resource group recovery on node startup is different for non-concurrent and concurrent resource groups:

- If the starting node fails to activate a *non-concurrent resource group* that is in the ERROR state, the resource group continues to selectively fall over to another node in the nodelist, if a node is available. In this case, HACMP uses selective fallover. The fallover action continues until all available nodes in the nodelist have been tried.
- If the starting node fails to activate a *concurrent resource group* that is in the ERROR state, the concurrent resource group is left in the ERROR state.

Note: HACMP 5.2 and up *automatically* attempts to activate resource groups in the ERROR state on the node during a **node_up** event. You cannot disable this functionality. If an attempt to recover a resource group in the ERROR state on a joining node is made but the resource group acquisition on the node fails, the non-concurrent resource groups falls over to the next node in the nodelist, if one is available. If the concurrent resource group acquisition fails, the resource group remains in the ERROR state.

Handling of Resource Groups Configured with IPAT via IP Aliases

When you configure your HACMP cluster, you define certain IP labels/IP addresses (service addresses) to be kept highly available. These service addresses are typically the IP address used by clients to access the server application. HACMP keeps the IP address available to clients by moving the address between different network interfaces.

With IPAT via IP Replacement there is a strict one-to-one mapping between IP address and network interface: Only one address can be placed on an interface at a given time. When HACMP moves the service IP address, it first removes the base or boot address from the interface, then puts the service address on the interface.

This one-to-one relationship between addresses and interfaces requires physical cluster hardware—a local interface and a backup interface—for every highly available service address. This also limits the configuration options available for the HACMP cluster since resource group placement is bound to the availability of physical resources and recovery is *not* possible once all the physical resources are consumed.

IP aliasing is a function of the TCP/IP stack where multiple IP addresses can be added to the same physical interface. When HACMP uses IP aliases for recovery, the base address for the interface does *not* change. HACMP recovers the service address by adding it as a second or alias address on the same interface. If you are using IPAT via IP Aliases, networks do *not* have the strict one-to-one relationship between interfaces and addresses. A single physical interface can host or back up multiple service addresses. This greatly improves the configuration flexibility and fallover options by requiring fewer physical resources to serve as backup. IPAT via IP Aliases is also faster as there are fewer commands required when moving addresses.

Do *not* mix service IP labels that can be used in IPAT via Aliases with those that cannot in the same resource group. This restriction is enforced during verification of cluster resources.

To control the placement of the service IP label aliases on the cluster node physical network interface cards, you can configure a distribution preference for the aliases of the service IP labels that are placed under HACMP control. See the section Distribution Preference for Service IP Label Aliases: Overview in Chapter 4: Configuring HACMP Cluster Topology and Resources (Extended).

Resource Group Behavior when Using IPAT via IP Aliases

Networks configured to use IPAT via Aliases have service addresses (the IP address HACMP is to keep highly available) and base interfaces (the network interfaces to use to host the service address). There are no “standby” interfaces. When multiple interfaces are on the same network and the same node, they are all defined as boot.

With IPAT via IP Aliases, the service address is added as an alias address on an available boot interface. This applies to the node where the resource group is first acquired, as well as to the node(s) that might acquire it later. When the resource group is released, the service address is removed from the interface, but this does *not* alter the base or boot address on the interface.

The mechanics of IP address takeover on a network using aliases works the same way for all non-concurrent resource groups. While the mechanics are identical, IPAT via IP aliases does affect the initial startup and fallover placement of the resource group.

The aliased service IP labels are distributed across all available boot interfaces. To facilitate even distribution of labels across all available IP interface cards, HACMP sorts all available interfaces by state and then by the number of aliased addresses already placed on the interface, and places the aliased labels accordingly. Note that this distribution is only done at fallover time, HACMP makes no attempt to redistribute the labels later, if another interface becomes active.

Note: If you want HACMP to activate only a *specific* resource group on a node during startup among multiple resource groups that could potentially be acquired on this node, we recommend that you use the startup policy Online Using Node Distribution Policy.

Resource Group Placement on Cluster Startup

The presence of a service IP label defined on an network configured for IPAT via aliases in the resource group does *not* alter the resource group's placement policy on *initial cluster startup*. That is, on initial cluster startup, the non-concurrent resource group configured with IPAT via IP Aliases is placed according to the defined startup policy.

On the subsequent cluster startup, HACMP moves the resource group containing the service IP label onto a node with a boot interface that:

- Is up
- Has a different subnet than the IP label that is being moved.

In addition, HACMP follows these rules:

- If multiple boot interfaces are found that are up and have different subnets, then HACMP moves the resource group onto the one that comes first in the alphabetically-sorted list of network interfaces configured on the node.
- If the resource group uses Online Using Node Distribution Policy startup, the resource group is placed on a node that does *not* host another resource group.

Resource Group Placement on Fallover

On fallover, if you have configured resource groups that contain aliased service IP labels, this allows having more than one non-concurrent resource group on the same node. Therefore, more than one resource group can be serviced by a node with a single physical interface.

On fallover, HACMP moves the resource group containing the service IP label onto a node with a boot interface which:

- Is up
- Has a different subnet
  - Is preferably *not* hosting another service label (if available).
  - Comes first in the alphabetically-sorted list of network interfaces in the network configuration.
  - If multiple boot interfaces are found that are up, have different subnets, and are *not* hosting another service label, then HACMP moves the resource group onto the one that comes first in the alphabetically-sorted list of network interfaces configured on the node.

The key advantage of having resource groups configured with IPAT via IP Aliases is that, on fallover, more than one resource group can be serviced by a node with a single physical interface.

Examples of Location Dependency and Resource Group Behavior

This appendix contains scenarios that illustrate how location dependent resource groups are processed at startup and also how they are processed for various failure scenarios.

Publishing Model with Same Node and Different Nodes Dependencies

The XYZ Publishing company follows a business continuity model that involves prioritizing the different platforms used to develop the web content. XYZ uses location dependency policies to keep some resource groups strictly on separate nodes and others together on the same node.

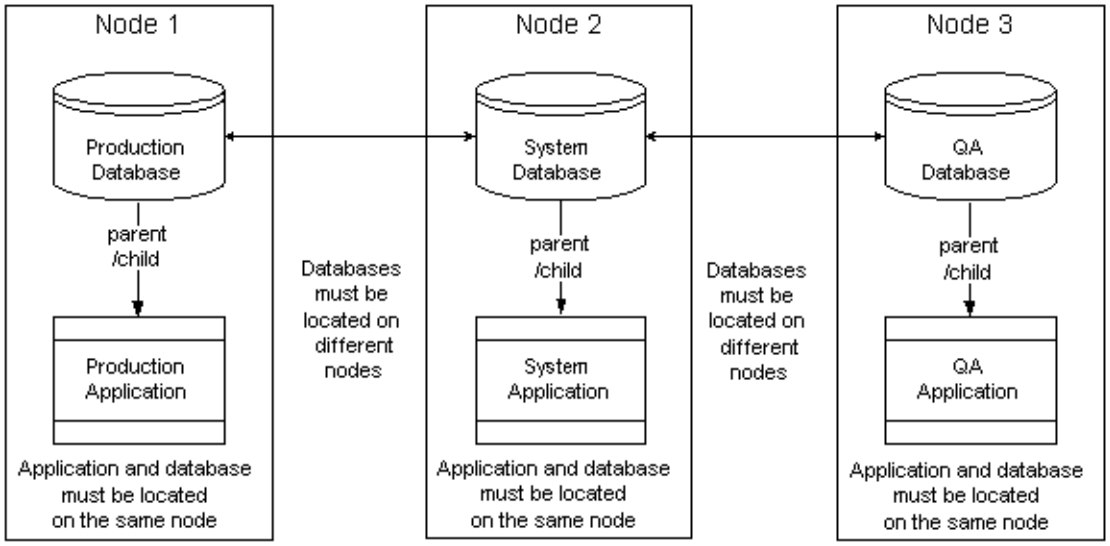
The Production database (PDB) and Production application (Papp) are hosted on the same node to facilitate maintenance (and perhaps the highest priority node for these resource groups has the most memory or faster processor). It also makes sense to set up a parent/child relation between them, since the application depends on the database. The database must be online for the application to function.The same conditions are true for the System Database (SDB) and the System application (Sapp) and for the QA Database (QADB) and the QA application (QAapp).

Since keeping the production database and application running is the highest priority, it makes sense to configure the cluster so that the three database resource groups stay on different nodes (make them an Online On Different Nodes dependency set), and assign the PDB resource group with the **high** priority. The SDB is the **Intermediate** priority and the QADB is the **low** priority.

The databases and their related applications are each configured to belong to an Online On Same Node dependency set.

HACMP handles these groups somewhat differently depending on how you configure startup, fallover, and fallback policies. It makes sense to have the participating nodelists differ for each database and application set to facilitate keeping these resource groups on the preferred nodes.

The figure below shows the basic configuration of the three nodes and six resource groups.



Publishing Model with Parent/Child and Location Dependencies

Resource Group Policies: Online on First Available Node

For the following use case discussions, all six resource groups have the following policies:

- Startup Policy: Online On First Available Node
- Fallover Policy: Fallover to Next Priority Node
- Fallback Policy: Never Fallback

Participating Nodes	Location Dependency	Parent/Child Dependency
• PApp: 1, 2, 3 • PDB: 1, 2, 3 • SApp: 2, 3 • SDB: 2, 3 • QAApp: 3 • QADB: 3	Online On The Same Node Dependent Groups: • PApp with PDB • SApp with SDB • QAApp with QADB Online On Different Nodes Dependent set: [PDB SDB QADB] Priority: PDB > SDB > QADB	• PApp (child) depends on PDB (parent) • SApp (child) depends on SDB (parent) • QAApp (child) depends on QADB (parent)

Use Case 1: Start Nodes in Numerical Order (Node1 First)

Starting the nodes in numerical order we expect the Production resource groups to come online on Node 1, the System resource groups to come online on Node 2, and the QA resource groups to come online on Node 3. There is no contention.

Node 1 is the highest priority node for resource groups PDB and PApp. The parent/child dependency dictates that PDB must be brought online prior to processing PApp. Therefore, HACMP processes the **rg_move** event to acquire PDB first and then it acquires PApp.

Node 1 is *not* in the nodelist for any other groups. Even if it were, the Online on Different Nodes dependency would disallow any lower priority groups from coming online on this node.

Consolidated View of Start Node Sequence: 1, 2, 3

Step	Node 1	Node 2	Node 3
Start Node 1	PApp: **ONLINE** PDB: **ONLINE**	PApp: PDB: SApp: SDB:	PApp: PDB: SApp: SDB: QAApp: QADB:
Start Node 2	PApp: **ONLINE** PDB: **ONLINE**	PApp: OFFLINE PDB: OFFLINE SApp: **ONLINE** SDB: **ONLINE**	PApp: PDB: SApp: SDB: QAApp: QADB:
Start Node 3	PApp: **ONLINE** PDB: **ONLINE**	PApp: OFFLINE PDB: OFFLINE SApp: **ONLINE** SDB: **ONLINE**	PApp: OFFLINE PDB: OFFLINE SApp: OFFLINE SDB: OFFLINE QAApp: **ONLINE** QADB: **ONLINE**

Use Case 2: Start Nodes Out of Order (Node 3)

Precondition: Resource groups are offline, all nodes are offline				
Step/ Qualifier	**Action**	**Node 1**	**Node 2**	**Node 3**
1	Start Node 3			
2				Acquire PDB
3				Acquire PApp
Post-condition/ Resource group states		PDB PApp	PDB: Papp SDB SApp	PApp: **ONLINE** PDB: **ONLINE** SApp: ERROR SDB: ERROR QAApp: ERROR QADB: ERROR

Node 3 is the lowest priority node for PDB and PApp, as well as for SDB and SApp. Node 3 is the highest priority node for QADB and QAApp. However, the PDB/PApp pair has the highest priority due to the Online On Different Nodes Dependency. Therefore, HACMP will acquire and start PDB on Node 3 and then process its child PApp. The other resource groups will go to the ERROR state based on the rule—these resource groups could have been brought online on Node 3 but were *not* acquired due to the Online On Different Nodes Dependency policy.

Use Case 2 Continued: Start Nodes Out of Order (Node 2)

Precondition: Node 3 is up; cluster and group states as at the end of the previous table.				
Step/Qualifier	**Action**	**Node 1**	**Node 2**	**Node 3**
1	Start Node 2			
2				Release PApp
3				Release PDB
4			Acquire PDB	Acquire SDB
5			Acquire PApp	Acquire SApp
Post-condition / Resource group states		PApp: PDB:	PApp: **ONLINE** PDB: **ONLINE** SApp: OFFLINE SDB: OFFLINE	PApp: OFFLINE PDB: OFFLINE SApp: **ONLINE** SDB: **ONLINE** QAApp: ERROR QADB: ERROR

Node 2 is the highest priority node for the SDB and SApp resource groups. However, the higher priority resource group in the Online On Different Nodes Dependency set is PDB. Therefore, PDB will fall over to this joining node while SDB and SApp will be acquired and started on Node 3. HACMP moves Papp to the same node with PDB because these two resource groups belong to an Online on Same Node dependency set. QA PDB is lower priority than SDB, so it stays in the ERROR state along with QAapp.

When Node 1 comes up, PDB and Papp will fall over to Node 1, SDB and Sapp will fall over to Node 2, and the QA resource groups will be acquired and started on Node 3.

Consolidated View of Start Nodes Out of Order: Sequence 3, 2, 1

Step	Node 1	Node 2	Node 3
Start Node 3	PApp: PDB:	PApp: PDB: SApp: SDB:	PApp: **ONLINE** PDB: **ONLINE** SApp: ERROR SDB: ERROR QAApp: ERROR QADB: ERROR
Start Node 2	PApp: PDB:	PApp: **ONLINE** PDB: **ONLINE** SApp: OFFLINE SDB: OFFLINE	PApp: OFFLINE PDB: OFFLINE SApp: **ONLINE** SDB: **ONLINE** QAApp: ERROR QADB: ERROR
Start Node 1	PApp: **ONLINE** PDB: **ONLINE**	PApp: OFFLINE PDB: OFFLINE SApp: **ONLINE** SDB: **ONLINE**	PApp: OFFLINE PDB: OFFLINE SApp: OFFLINE SDB: OFFLINE QAApp: **ONLINE** QADB: **ONLINE**

Use Case 3: Fallover of Resource Groups due to Node Failure

Precondition: All nodes are ONLINE. Resource groups PDB and PApp are online on Node 1, SDB and SApp are online on Node 2, QAApp and QADB are online on Node 3					
Step/ Qualifier	**Action**	**Node 1**	**Node 2**	**Node 3**	**Comments**
1	Node 1 crash.		Release SApp	Release QAApp	
2			Release SDB	Release QADB	QAApp and QDB go to ERROR state.
3			Acquire PDB	Acquire SDB	
4			Acquire PApp	Acquire SApp	
Post-condition /Resource Group states		PApp: PDB:	PApp: **ONLINE** PDB: **ONLINE** SApp: ERROR SDB: ERROR	PApp: OFFLINE PDB: OFFLINE SApp: **ONLINE** SDB: **ONLINE** QAApp: ERROR QADB: ERROR	

When Node 1 fails, HACMP releases SApp and SDB and QADB and QAapp and moves the highest priority resource group PDB and its Same Node dependency partner and child PApp to Node 2 and likewise moves the System groups to Node 3. The QA groups are left with nowhere to go; they go into ERROR state).

Use Case 4: Fallover of Resource Groups—Network Goes Down During Fallover

Precondition: All nodes are ONLINE. Resource groups PDB and PApp are online on Node 1, SDB and SApp are online on Node 2, QAApp and QADB are online on Node 3. All applications use the app_network,					
Step/ Qualifier	**Action**	**Node 1**	**Node 2**	**Node 3**	**Comments**
1	Node 1 crash.		Release SApp	Release QAApp	
2			Release SDB	Release QADB	
3			Acquire PDB	Acquire SDB	QADB goes into ERROR state
4	app_network down Node 2				app_network failure.
5			Acquire PApp	Acquire SApp	PApp and SApp go to the ERROR state (network *not* available)
6			resource_state_ change event	resource_state_ change event	Triggers **rg_move** events
7			Release PDB	Release SApp	
8				Release SDB	
9			Acquire SDB	Acquire PDB	
10			Acquire SApp	Acquire PApp	
Post-condition/ Resource group states		PApp: PDB:	PApp: OFFLINE PDB: OFFLINE SApp: ERROR SDB: **ONLINE**	PApp: **ONLINE** PDB: **ONLINE** SApp: ERROR SDB: ERROR QAApp: ERROR QADB: ERROR	

In step 5, PApp goes to the ERROR state directly, instead of going through the acquisition phase since the Cluster Manager knows that the network required for PApp on Node 2 is currently down. This is in contrast to an acquisition failure.

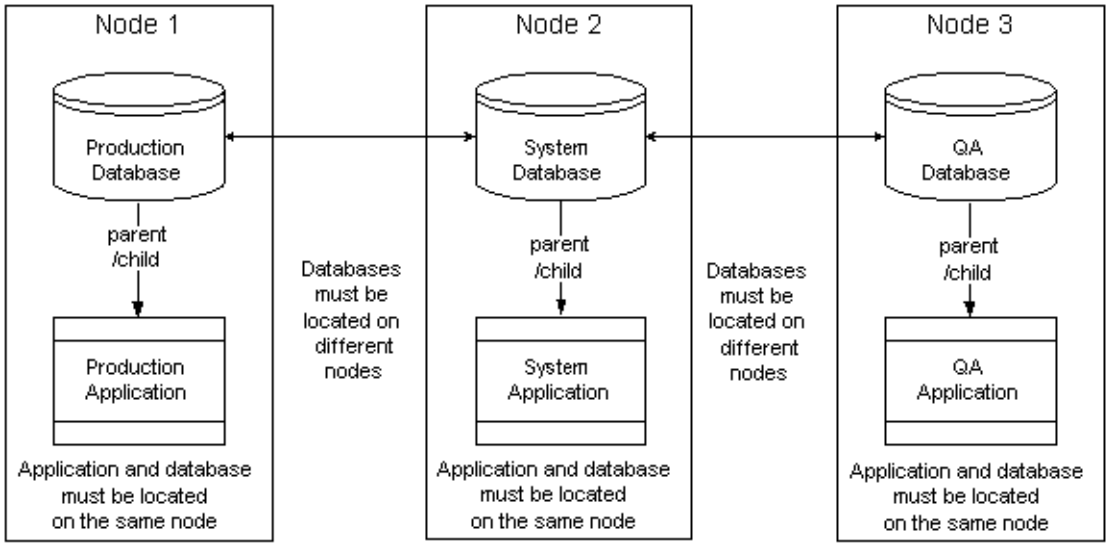
In step 6, the event queue gets a resource_state_change event on the queue, which gets voted and queues additional ACQUIRE/RELEASE events.

In steps 7 & 8: SApp goes to the ERROR state due to network failure.

Publishing Model—Alternate Configuration

This model consists of three pairs of parent and child resource groups (total of six resource groups) and three nodes in the HACMP cluster. The applications (PApp, SApp and QAApp) are Online On The Same Node with their corresponding databases (PDB, SDB and QADB.) All databases (which are parent resource groups also) are Online On Different Nodes from the other databases.

The only difference between the original Publishing Model configuration and this alternate Publishing Model is the resource group's startup preferences. This section uses the **Online On Home Node Only** startup policy whereas the original Publishing Model configuration uses **Online On First Available Node** as the startup policy for the resource groups.



Alternate Publishing Model: Startup on Home Node Only

Resource Group Policies: Online on Home Node Only

All six resource groups have the following policies:

- Startup Policy: Online On Home Node Only—this is different from the previous set of use cases.
- Fallover Policy: Fallover to Next priority node
- Fallback Policy: Never Fallback

•

Participating Nodes	Location Dependency	Parent/Child Dependency
PApp: 1, 2, 3 PDB: 1, 2, 3 SApp: 2, 3 SDB: 2, 3 QAApp: 3 QADB: 3	Online On The Same Node Dependent Groups: • PApp is with PDB • SApp is with SDB • QAApp is with QADB Online On Different Nodes Dependent set: [PDB SDB QADB] Priority: PDB>SDB>QADB	PApp (child) depends on PDB (parent) SApp (child) depends on SDB (parent) QAApp (child) depends on QADB (parent)

Use Case 1: Start Lowest Priority Node (Node 3)

Precondition: All resource groups are offline, all nodes are offline				
Step/ Qualifier	**Action**	**Node 1**	**Node 2**	**Node 3**
1	Start Node 3			
2				Acquire QADB
3				Acquire QAApp
Post-condition /Resource group states		PApp: PDB:	Papp PDB: SApp: SDB:	PApp: OFFLINE PDB: OFFLINE SApp: OFFLINE SDB: OFFLINE QAApp: **ONLINE** QADB: **ONLINE**

Node 3 is the home node for resource groups QAApp and QADB. Although PDB and PApp have higher priority as defined by the Online On Different Nodes Dependency set, during cluster startup the startup policy allows only the QAApp and QADB resource groups to come online on Node 3. Therefore, the higher priority resource groups remain in the OFFLINE state at startup.

Use Case 2: Start Second Node (Node 2)

Precondition: Node 3 is up; cluster and group states as at the end of the previous use case.				
Step/Qualifier	**Action**	**Node 1**	**Node 2**	**Node 3**
1	Start Node 2			
2			Acquire SDB	
3			Acquire SApp	
Post-condition / Resource group states		PApp: PDB:	PApp: OFFLINE PDB: OFFLINE SApp: **ONLINE** SDB: **ONLINE**	PApp: OFFLINE PDB: OFFLINE SApp: OFFLINE SDB: OFFLINE QAApp: **ONLINE** QADB: **ONLINE**

Node 2 is the highest priority node for SDB and SApp resource groups. Since the startup policy for the resource groups is Online On Home Node, these resource groups will be started even though PDB and PApp are the highest priority resource groups.

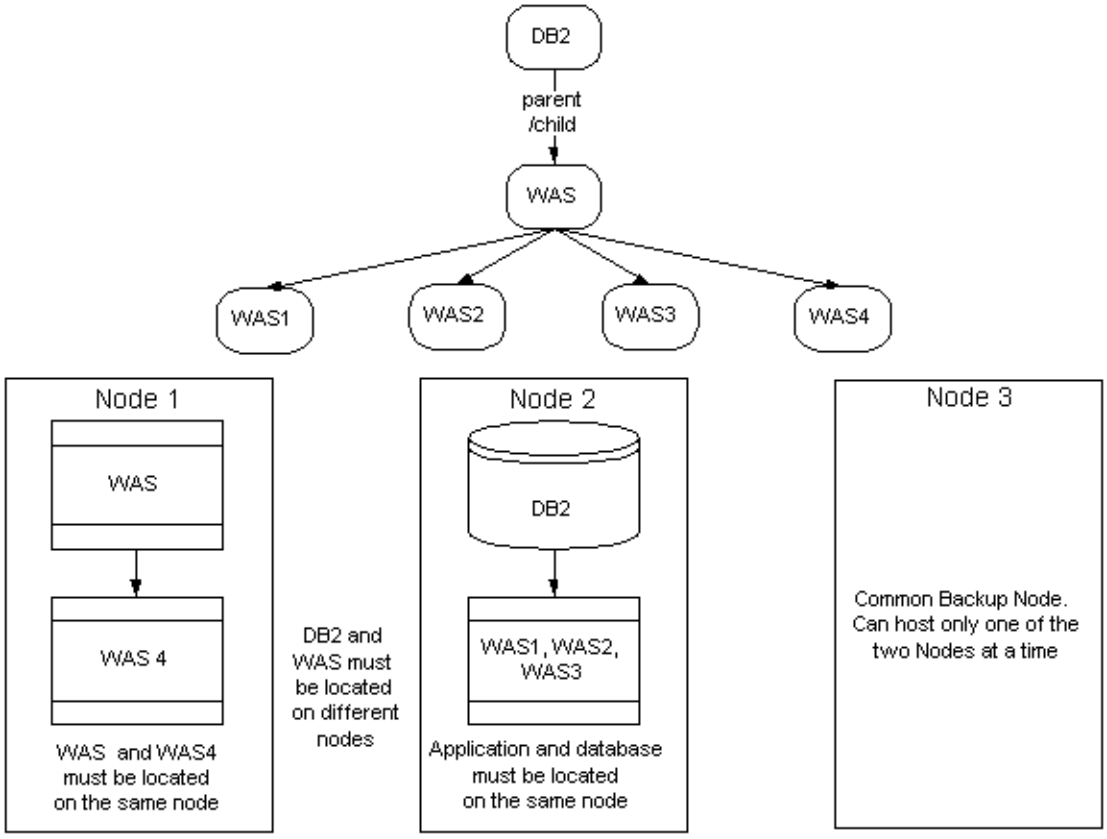
Consolidated View of Start Node Sequence: 3, 2, 1

Step	Node 1	Node 2	Node 3
Start Node 3	PApp: PDB:	PApp: PDB: SApp: SDB:	PApp: OFFLINE PDB: OFFLINE SApp: OFFLINE SDB: OFFLINE QAApp: **ONLINE** QADB: **ONLINE**
Start Node 2	PApp: PDB:	PApp: **ONLINE** PDB: **ONLINE** SApp: OFFLINE SDB: OFFLINE	PApp: OFFLINE PDB: OFFLINE SApp: OFFLINE SDB: OFFLINE QAApp: **ONLINE** QADB: **ONLINE**
Start Node 1	PApp: **ONLINE** PDB: **ONLIN**E	PApp: OFFLINE PDB: OFFLINE SApp: **ONLINE** SDB: **ONLINE**	PApp: OFFLINE PDB: OFFLINE SApp: OFFLINE SDB: OFFLINE QAApp: **ONLINE** QADB: **ONLINE**

WAS/DB2 Cluster Model and Use Cases

This model contains a DB2 database, a WAS application that depends on DB2 and four WebSphere Applications. The parent/child dependency for this model is that DB2 should be available prior to activating the WAS and WebSphere (WS#) applications depend on the availability of WAS.

The location dependency of the resource groups is that DB2 and WAS should *not* be activated on the same node and WAS is Online On The Same Node Dependent with WS4 (see figure) and DB2 is Online On The Same Node with WS1, WS2 and WS3. The location dependency for the WS# is purely artificial in this example. However, this is a configuration where one of the nodes is fine-tuned for DB2 (hence will be the highest priority node for DB2) and the other one is fine-tuned for WAS. They both have a common backup node, which can host only one of the two groups at a time.



WAS/DB2 Cluster with Location and Parent/Child Dependencies

Resource Group Policies

All resource groups have the following policies:

- Startup Policy: Online On First Available Node
- Fallover Policy: Fallover to Next priority node
- Fallback Policy: Never Fallback

Participating Nodes	Location Dependency	Parent/Child Dependency
DB2 [2, 3] WS1 [2, 3] WS2 [2, 3] WS3 [2, 3] WS4 [1, 3] WAS [1, 3]	Online On The Same Node Dependent Groups: 1. DB2, WS1, WS2, WS3 2. WAS, WS4 Online On Different Nodes Dependent Groups: • DB2, WAS	1. WS1, WS2, WS3 and WS4 (children) depend on WAS (parent) 2. WAS (child) depends on DB2 (parent)

Use Case 1: Start First Node (Node 1)

Precondition: All resource groups are offline, all nodes are offline				
Step/ Qualifier	**Action**	**Node 1**	**Node 2**	**Node 3**
1	Start Node 1			Parent/child dependency *not* met.
2		WAS: ERROR		
3		WS4: ERROR		
Post-condition / Resource group states		WAS: ERROR WS4: ERROR	DB2: WS1: WS2: WS3:	WAS: DB2: WS1: WS2: WS3: WS4

WAS and WS4 could have started on Node 1, but the parent resource group DB2 is still in the offline state. Therefore, WAS and WS4 are put in the ERROR state.

Use Case 2: Start Second Node (Node 2)

Precondition: Cluster state as in the post-condition from the above use case.				
Step/ Qualifier	**Action**	**Node 1**	**Node 2**	**Node 3**
1	Start Node 2			
2			Acquire DB2	
3		Acquire WAS		
4		Acquire WS4	Acquire WS1, WS2, WS3	
Post-condition / Resource group states		WAS:**ONLINE** WS4: **ONLINE**	DB2: **ONLINE** WS1: **ONLINE** WS2: **ONLINE** WS3: **ONLINE**	WAS: DB2: WS1: WS2: WS3: WS4:

Node 2 starts DB2 (the parent RG), which in turn triggers processing of WAS (child of DB2). Finally all the grandchildren are started on their respective nodes.

Consolidated View of Start Node Sequence 1, 2, 3

Step	Node 1	Node 2	Node 3
Start node 1	WAS: ERROR WS4: ERROR	DB2: WS1: WS2: WS3:	WAS: DB2: WS1: WS2: WS3: WS4:
Start node 2	WAS: **ONLINE** WS4: **ONLINE**	DB2: **ONLINE** WS1: **ONLINE** WS2: **ONLINE** WS3: **ONLINE**	WAS: DB2: WS1: WS2: WS3: WS4:
Start node 3	WAS: **ONLINE** WS4: **ONLINE**	DB2: **ONLINE** WS1: **ONLINE** WS2: **ONLINE** WS3: **ONLINE**	WAS: OFFLINE DB2: OFFLINE WS1: OFFLINE WS2: OFFLINE WS3: OFFLINE WS4: OFFLINE

Use Case 3: Start Nodes Out of Order (Node 3)

Precondition: All cluster nodes and resource groups are in the offline state.					
Step/ Qualifier	**Action**	**Node 1**	**Node 2**	**Node 3**	**Comments**
1	Start Node 3				
2				Acquire DB2	
Post-condition / Resource group states		WAS: DB2: WS4	WS1: WS2: WS3:	WAS: ERROR DB2: **ONLINE** WS1: ERROR WS2: ERROR WS3: ERROR WS4: ERROR	

Node 3 is a participating node for all the resource groups. However, WAS and DB2 cannot coexist on the same node. DB2— being a parent—is started on Node 3, which means that WAS cannot be started on the same node. Since WAS is *not* online none of the children of WAS can come online on Node 3.

Use Case 4: Start Second Node Out of Order (Node 2)

Precondition: Cluster and RG states as at the end of the previous use case.				
Step/ Qualifier	**Action**	**Node 1**	**Node 2**	**Node 3**
1	Start Node 2			
2				Release DB2
3			Acquire DB2	
4				Acquire WAS
			Acquire WS1, WS2, WS3	Acquire WS4
Post-condition / Resource group states		WAS: WS4	DB2: **ONLINE** WS1: **ONLINE** WS2: **ONLINE** WS3: **ONLINE**	WAS: **ONLINE** DB2: OFFLINE WS1: OFFLINE WS2: OFFLINE WS3: OFFLINE WS4: **ONLINE**

Node 2 is the higher priority node for DB2. Therefore DB2 falls back to Node 2 and WAS (Online On Different Nodes Dependency set) can now be acquired on Node 3.

Use Case 5: Start Third Node (Node)1

Precondition: Cluster and RG states as at the end of the previous use case.				
Step/ Qualifier	**Action**	**Node 1**	**Node 2**	**Node 3**
1	Start Node 1			
2			Release WS1, WS2, and WS3	Release WS4
3		Acquire WAS		
4		Acquire WS4		
5			Acquire WS1, WS2 and WS3	
Post-condition / Resource group states		WAS: **ONLINE** WS4: **ONLINE**	DB2: **ONLINE** WS1: **ONLINE** WS2: **ONLINE** WS3: **ONLINE**	WAS: OFFLINE DB2: OFFLINE WS1: OFFLINE WS2: OFFLINE WS3: OFFLINE WS4: OFFLINE

All groups are now online.

Consolidated View of Start Node Sequence: 3, 2, 1

Step	Node 1	Node 2	Node 3
Start Node 3	WAS: WS4:	DB2: WS1: WS2: WS3:	WAS: ERROR DB2: **ONLINE** WS1: ERROR WS2: ERROR WS3: ERROR WS4: ERROR
Start Node 2	WAS: WS4:	DB2: **ONLINE** WS1: **ONLINE** WS2: **ONLINE** WS3: **ONLINE**	WAS: **ONLINE** DB2: OFFLINE WS1: OFFLINE WS2: OFFLINE WS3: OFFLINE WS4: **ONLINE**
Start Node 1	WAS: **ONLINE** WS4: **ONLINE**	DB2: **ONLINE** WS1: **ONLINE** WS2: **ONLINE** WS3: **ONLINE**	WAS: OFFLINE DB2: OFFLINE WS1: OFFLINE WS2: OFFLINE WS3: OFFLINE WS4: OFFLINE

Use Case 6: Acquisition Failure Example

Precondition: Node 1 is offline and all resource groups are **ONLINE** on Nodes 2 and 3.					
		Node 1	**Node 2**	**Node 3**	
		WAS: WS4:	DB2: **ONLINE** WS1: **ONLINE** WS2: **ONLINE** WS3:**ONLINE**	WAS: **ONLINE** DB2: OFFLINE WS1: OFFLINE WS2: OFFLINE WS3: OFFLINE WS4: **ONLINE**	
Step/ Qualifier	**Action**	**Node 1**	**Node 2**	**Node 3**	**Comments**
1	Node_up 1				
2			Release WS1 WS2 WS3 Release	Release WS4	
3				Release WAS	
4		Acquire WAS			Acquisition Failure for WAS
5		rg_move WAS			Normal rg_move event
6				Acquire WAS	
7			Acquire WS1 WS2 WS3 Acquire	Acquire WS4	
Post condition/ Resource group states		WAS:OFFLINE WS4: OFFLINE	DB2: **ONLINE** WS1: **ONLINE** WS2: **ONLINE** WS3: **ONLINE**	WAS: **ONLINE** DB2: OFFLINE WS1: OFFLINE WS2: OFFLINE WS3: OFFLINE WS4: **ONLINE**	

As Node 1 joins the cluster, WAS attempts to fallback but gets the acquisition failure. The acquisition failure launches a **resource_state_change** event; this triggers an **rg_move** event, which moves WAS to its original node.

Use Case 7: Resource Group Move Due to Local Network Down (DB2 failure)

Precondition: All resource groups are ONLINE and all nodes are ONLINE.					
		Node 1	**Node 2**	**Node 3**	
		WAS: WS4:	DB2: **ONLINE** WS1: **ONLINE** WS2: **ONLINE** WS3:**ONLINE**	WAS: **ONLINE** DB2: OFFLINE WS1: OFFLINE WS2: OFFLINE WS3: OFFLINE WS4: **ONLINE**	
Step/ Qualifier	**Action**	**Node 1**	**Node 2**	**Node 3**	**Comments**
1	Local network failure for DB2				
2		Release WS4	Release WS1 WS2 WS3		
3		Release WAS			
4			Release DB2		
5				Acquire DB2	
6		Acquire WAS			
7		Acquire WS4		Acquire WS1 WS2 WS3	
Post condition/ Resource group states		WAS: OFFLINE WS4: OFFLINE	DB2: **ONLINE** WS1: **ONLINE** WS2: **ONLINE** WS3: **ONLINE**	WAS: **ONLINE** DB2: OFFLINE WS1: OFFLINE WS2: OFFLINE WS3: OFFLINE WS4: **ONLINE**	

The local network down for DB2 (the parent group) results in release and acquire of the entire stack of the resource groups.

Replicated WAS-Solution Model Use Cases

This model contains a DB2 database and a WAS application that depends on DB2; DB2 in turn depends on an LDAP directory server. Note that DB2 normally does *not* depend on LDAP, but this scenario is used for our discussion.

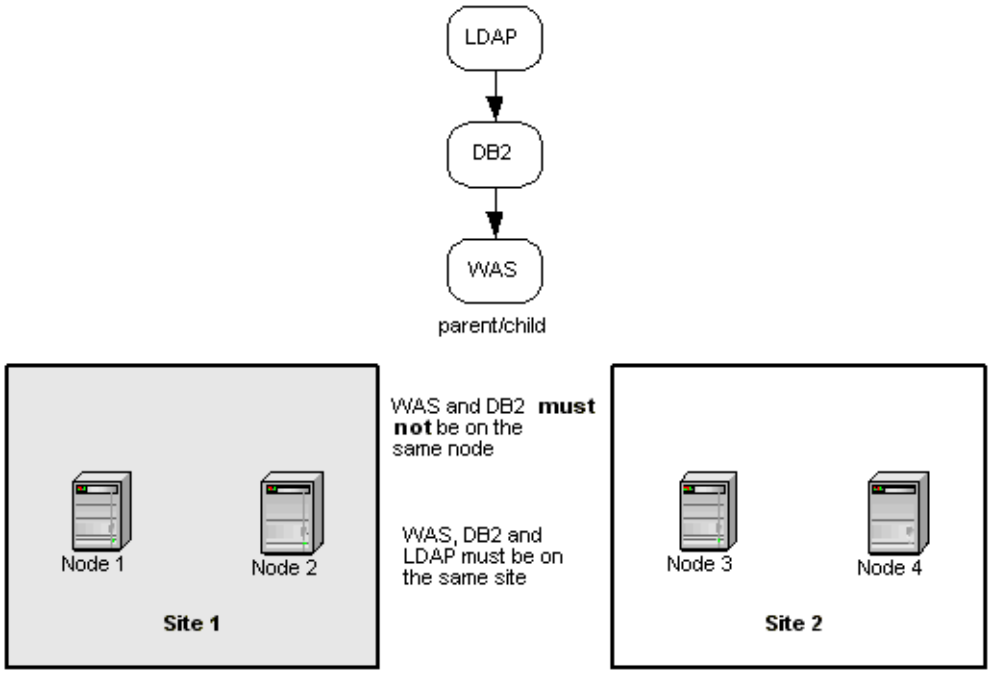
The parent/child dependency for this model is that LDAP should be available before activating DB2 and DB2 should be available prior to activating the WAS.

The location dependency of the resource groups is that DB2 and WAS should *not* be activated on the same node. LDAP can coexist with either of the applications.

The model has a backup site as well, and the constraint that, for performance reasons, all three resource groups should reside on the same site at any given time. In other words, the three resource groups involved have the dependency Online On The Same Site.

All resource groups are of equal priority. There is an indirect priority implied by the parent/child dependency configuration, *not* something the administrator explicitly specified.

Replicated WAS-Solution, Configuration I



Replicated WAS Solution with Parent/Child, Different Nodes and Same Site Dependencies

Resource Group Policies

All resource groups have the following policies:

- Startup Policy: Online On First Available Node
- Fallover Policy: Fallover to Next priority node
- Fallback Policy: Never Fallback
- Site Policy: Prefer Primary Site

Participating Nodes	Location Dependency	Parent/Child Dependency
DB2 [1, 2, 3, 4] LDAP [1, 2, 3, 4] WAS [1, 2, 3, 4]	Online On The Same Site Dependent Groups: • DB2, LDAP, WAS Online On Different Nodes Dependent Groups: • DB2, WAS	1. DB2 (child) depends on LDAP (parent.) 2. WAS (child) depends on DB2 (parent.)

Use Case 1: Start First Node (Node 1)

Precondition: All resource groups are offline, all nodes are offline						
Step/ Qualifier	**Action**	**Node 1**	**Node 2**	**Node 3**	**Node 4**	**Comments**
1	Start Node 1					
2		LDAP Primary instance is acquired				
3		DB2 Primary Instance is acquired				
4		WAS goes into global ERROR state				Startup policy has been met, but no node available
Post-condition/ Resource group states		LDAP: **ONLINE** DB2: **ONLINE** WAS: ERROR				

According to the resource group startup preferences, all resource groups could have come online on Node 1. LDAP came online first, DB2 second, which was dictated by the resource group parent/child dependency configuration. Since WAS and DB2 are Online On Different Nodes dependent, WAS could *not* come online, therefore it went into global ERROR state.

WAS going into ERROR state at this point might *not* be intuitive at first.

Use Case 2: Start Second Node (Node 2)

Precondition: See Post-condition for Use Case 1						
Step/ Qualifier	**Action**	**Node 1**	**Node 2**	**Node 3**	**Node 4**	**Comments**
1	Start Node 2					
2			WAS Primary Instance acquired			
Post condition/ Resource group states		LDAP: **ONLINE** DB2: **ONLINE**	WAS: **ONLINE**			

WAS is already in a pre-event ERROR state and has all of its dependencies met. Its parent resource groups are ONLINE. Node 2 is a candidate node because it is in the same site as Node 1 (the node hosting DB2 and LDAP) therefore, WAS Primary instance is acquired.

Use Case 3: Start Third Node (Node 3)

Precondition: See Post-condition for previous use case						
Step/ Qualifier	**Action**	**Node 1**	**Node 2**	**Node 3**	**Node 4**	**Comments**
1	Start Node 3					All secondary instances are acquired in parallel
2					Acquire: DB2, LDAP, and WAS Secondary	
Post condition/ Resource group states		LDAP: **ONLINE** DB2: **ONLINE**	WAS: **ONLINE**		LDAP: **Secondary** DB2: **Secondary** WAS: **Secondary**	

All secondary instances are acquired on the node that is started on the secondary site. Secondary instances do *not* repel each other; i.e., the Online On Different Nodes and Online On The Same Site policies that have been configured for the primary instances do *not* affect the behavior of the secondary instances. The secondary instances are acquired on the first node that joins the cluster.

Use Case 4: Start Fourth Node (Node 4)

Precondition: See Post-condition for previous use case						
Step/ Qualifier	**Action**	**Node 1**	**Node 2**	**Node 3**	**Node 4**	**Comments**
1	Start Node 4				no action	
2					Acquire: DB2, LDAP, and WAS Secondary	
Post-condition / Resource group states		LDAP: **ONLINE** DB2: **ONLINE**	WAS: **ONLINE**		LDAP: **Secondary** DB2: **Secondary** WAS **Secondary**	

No action is taken on any of the instances on any of the resource groups.

Use Case 5: User-requested rg_move Entire Stack to Opposite Site

Precondition: See Post-condition for previous use case						
Step/ Qualifier	**Action**	**Node 1**	**Node 2**	**Node 3**	**Node 4**	**Comments**
1	rg_move Same Site Dependent stack to Site 2					
2			Release WAS Primary			
3		Release DB2 Primary				
4		Release LDAP Primary				
5				Release: LDAP, DB2, WAS Secondary		
6		Acquire: LDAP, DB2, WAS Secondary				
7				Acquire LDAP Primary		
8				Acquire DB2 Primary		
9					Acquire WAS Primary	
Post condition/ Resource group states		LDAP: **Secondary** DB2: **Secondary** WAS **Secondary**			LDAP: **ONLINE** DB2: **ONLINE**	WAS: **ONLINE**

The primary instances are released prior to releasing the secondary instances, so that pending writes can be flushed to the secondary site.

During acquisition, the secondary instances are acquired prior to acquiring the primary instances, so that the primary instances can mirror as soon as they come online. HACMP is responsible for placing the primary instances on site_2, taking into consideration their parent/child and location dependencies. The secondary instances are placed on the opposite site.

The only constraint raised by resource behavior is that the primary instance MUST be released prior to releasing the secondary instance.

Use Case 6: DARE Change Nodelist of All Resource Groups

In this use case, we examine the steps used to handle the case when the node list of every resource group is changed, so that they are forced to move to the opposite site. The expected result, as well as the steps that need to be taken to get to that result, is similar to use case 5.

The node list in the old configuration is [1, 2, 3, 4] is changed to [4, 3, 2, 1]. Since the site policy is **Prefer Primary Site**, the DARE operation should move all resource groups to their new highest priority site (if at all possible), and distribute the resource groups on that site within the constraints of their various policies, such as parent/child and location dependency policies.

Precondition: See Post-condition for use case 4; all resource groups online on site1, and all nodes are available						
Step/ Qualifier	**Action**	**Node 1**	**Node 2**	**Node 3**	**Node 4**	**Comments**
1	DARE IN changed node list					
2			Release WAS Primary			
3		Release DB2 Primary				
4		Release LDAP Primary				
5				Release: LDAP, DB2, WAS Secondary		

Precondition: See Post-condition for use case 4; all resource groups online on site1, and all nodes are available						
Step/ Qualifier	**Action**	**Node 1**	**Node 2**	**Node 3**	**Node 4**	**Comments**
6		Acquire: LDAP, DB2, WAS Secondary				
					Acquire: LDAP	
					Acquire: DB2	
						Acquire WAS
Post-condition/ Resource group states		LDAP: **Secondary** DB2: **Secondary** WAS **Secondary**		LDAP: **ONLINE** DB2: **ONLINE**	WAS: **ONLINE**	

It is determined that all the resource groups can potentially be hosted on their new highest priority site; this is determined by taking into consideration the list of available nodes as well as the list of available interfaces. Then HACMP will try to move the entire stack to the opposite site.

The primary instances are released prior to releasing the secondary instances, so that pending writes can be flushed to the secondary site.

During acquisition, the secondary instances are acquired prior to acquiring the primary instances, so that the primary instances can mirror as soon as they come online. HACMP is responsible for placing the primary instances on site_2, taking into consideration their parent/child and location dependencies. The secondary instances are placed on available nodes at the opposite site.

Use Case 7: Node 4 fails

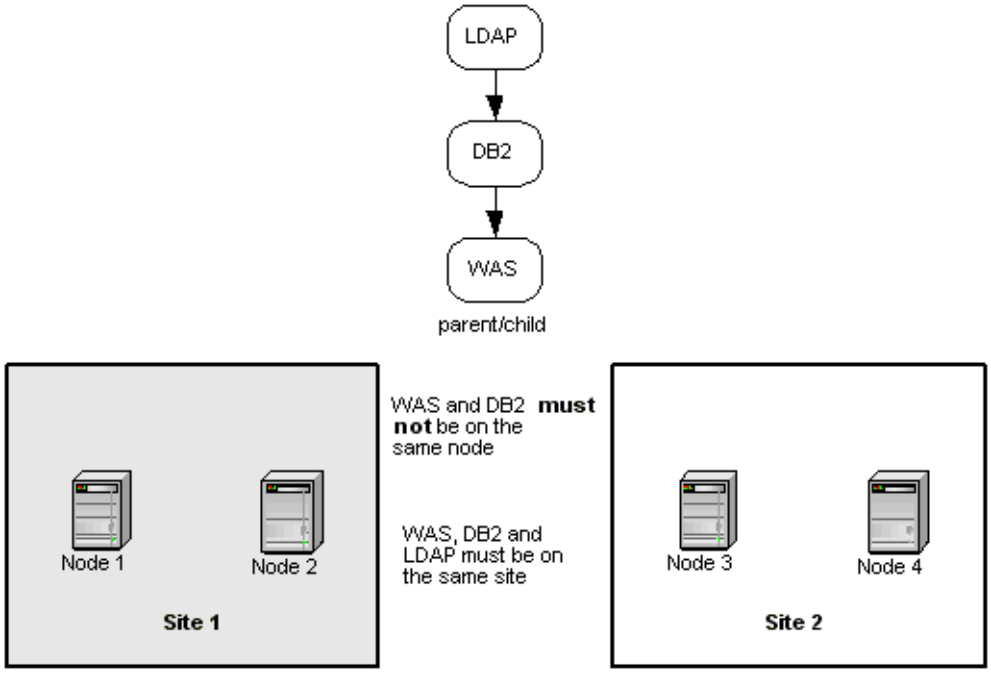
Precondition: See Post-condition for use case 5 in this section						
Step/ Qualifier	**Action**	**Node 1**	**Node 2**	**Node 3**	**Node 4**	**Comments**
1	Node 4 fails					
2					Release WAS Primary	
3				Release DB2 Primary		
4				Release LDAP Primary		
5		Release: LDAP, DB2, and WAS Secondary				
6				Acquire: LDAP, DB2, and WAS Secondary		
7		Acquire LDAP Primary				
8		Acquire DB2 Primary				
			Acquire WAS Primary			
Post-condition/ Resource group states		LDAP: **ONLINE** DB2: **ONLINE**	WAS: **ONLINE**		LDAP: **Secondary** DB2: **Secondary** WAS **Secondary**	

Since the node hosting WAS failed, the entire stack can be moved over to the opposite site—at least in this case. The primary instances are released prior to releasing the secondary instances, so that pending writes can be flushed to the secondary site.

During acquisition, the secondary instances are acquired prior to acquiring the primary instances, so that the primary instances can mirror as soon as they come online. HACMP is responsible for placing the primary instances on site_2, taking into consideration their parent/child and location dependencies. The secondary instances are placed on available nodes on the opposite site.

Issues are much the same as for the previous use case. This use case just reiterates the difficulties in handling the placement of the secondary instance, when the primary instance is *not* yet online.

Replicated WAS-Solution, Configuration II



Replicated WAS Solution II with Dependencies

Resource Group Policies

All resource groups have the following policies (note that Startup and Fallback policies changed from Configuration I):

- Startup Policy: Online On Home Node
- Fallover Policy: Fallover to Next priority node
- Fallback Policy: Fallback to Higher Priority Node
- Site Policy: Prefer Primary Site

Participating Nodes	Location Dependency	Parent/Child Dependency
DB2 [1, 2, 3, 4] LDAP [1, 2, 3, 4] WAS [2, 1, 3, 4]	Online On The Same Site Dependent Groups: • DB2, LDAP, WAS Online On Different Nodes Dependent Groups: • DB2, WAS	1. DB2 (child) depends on LDAP (parent.) 2. WAS (child) depends on DB2 (parent.)

Use Case 1: Start First Node (Node 2)

Precondition: All resource groups are offline, all nodes are offline						
Step/ Qualifier	**Action**	**Node 1**	**Node 2**	**Node 3**	**Node 4**	**Comments**
1	Start Node 2					
2			WAS goes into global ERROR state			Startup policy has been met, but parents *not* online
Post-condition / Resource group states			WAS: **ERROR**			

Since the WAS startup preference has been met, but its parent resource groups are *not* yet online, WAS goes into global ERROR state.

Use Case 2: Start Second Node (Node 4)

Precondition: See post-condition for the previous use-case						
Step/ Qualifier	**Action**	**Node 1**	**Node 2**	**Node 3**	**Node 4**	**Comments**
1	Start Node 4					
2					Acquire: DB2, LDAP, and WAS Secondary	Acquisition in parallel
Post-condition / Resource group states			WAS: **ERROR**		DB2: **Secondary** LDAP: **Secondary** WAS: **Secondary**	

The secondary instances are acquired as soon as possible. The startup preference, as well as parent/child and location dependencies, only applies to the primary instance. In a later use case we see that the fallback preference does *not* apply to the secondary instance either.

Use Case 3: Start Third Node (Node 3)

Precondition: See Post-condition for the previous use case.						
Step/ Qualifier	**Action**	**Node 1**	**Node 2**	**Node 3**	**Node 4**	**Comments**
1	Start Node 3					
2				No action		
Post-condition / Resource group states			WAS: ERROR		DB2: **Secondary** LDAP: **Secondary** WAS: **Secondary**	

The secondary instances do *not* fall back to the higher priority node. Fallback would mean an outage in mirroring.

Use Case 4: Bring Business Operation Online on Backup Site

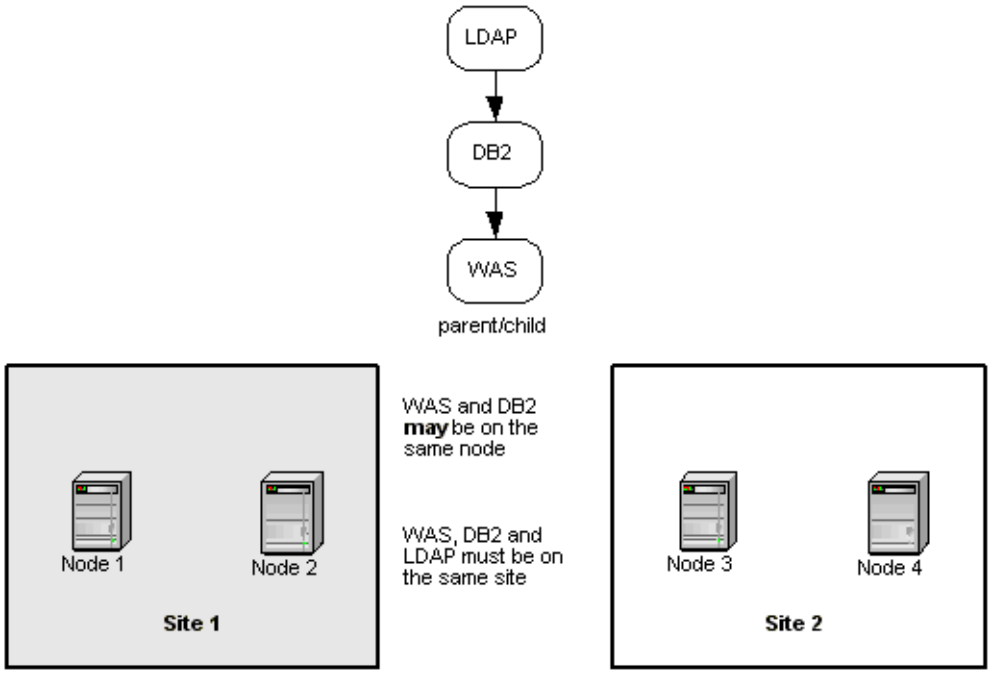
This use case illustrates what the administrator, as well as HACMP, has to do to bring business operations online on the secondary site. Node 1 on the primary site is *not* expected to ever join, all resource groups should be brought online on the secondary site, and since Node 2 is available we establish mirroring with the primary site.

Precondition: See post-condition for the previous use-case						
Step/ Qualifier	**Action**	**Node 1**	**Node 2**	**Node 3**	**Node 4**	**Comments**
1	rg_move online WAS on Node 4 /rg_move online entire stack on backup site					
2					Release DB2, WAS and LDAP Secondary	
			Acquire LDAP, DB2, and WAS Secondary			Parallel Operation
				Acquire LDAP Primary		
				Acquire DB2 Primary		
					Acquire WAS Primary	
Post-condition / Resource group states			DB2: **Secondary** LDAP: **Secondary** WAS: **Secondary**	DB2: **ONLINE** LDAP: **ONLINE** WAS: **ONLINE**	WAS: **ONLINE**	

As part of one user-requested **rg_move** operation, complex resource group movements have to happen in the correct order for both instances. The secondary instances are first brought offline on the site that will serve as the business operation site, then the secondary instances are brought online on the site that will serve as the backup site, after which the primary instances are placed on their new site, taking into consideration their parent/child and location dependency constraints.

If acquisition failures occur, the resource groups do *not* migrate back to their primary site, instead they go into ERROR state.

Replicated WAS-Solution, Configuration III



WAS Solution III - Concurrent Resource Group

This configuration is similar to the previous one, except that the DB2 resource group is a concurrent resource group and WAS and DB2 can come online on the same node.

Resource Group Policies

WAS and LDAP resource groups have the following policies:

- Startup Policy: Online On Home Node
- Fallover Policy: Fallover to Next priority node
- Fallback Policy: Fallback to Higher Priority Node
- Site Policy: Prefer Primary Site

DB2 has the following polices:

- Startup Policy: Online On All Available Nodes
- Fallover Policy: Bring Offline (On Error Node)
- Fallback Policy: Never Fallback

- Site Policy: Prefer Primary Site

Participating Nodes	Location Dependency	Parent/Child Dependency
DB2 [1, 2, 3, 4] LDAP [1, 2, 3, 4] WAS [2, 1, 3, 4]	Online On The Same Site Dependent Groups: 1. DB2 LDAP WAS 2. LDAP and DB2 should be on same node.	1. DB2 (child) depends on LDAP (parent.) 2. WAS (child) depends on DB2 (parent.)

Use Case 1: Start First Node (Node 2)

Precondition: All resource groups are offline, all nodes are offline.						
Step/ Qualifier	**Action**	**Node 1**	**Node 2**	**Node 3**	**Node 4**	**Comments**
1	Start Node 2		Acquire WAS			Goes to ERROR
2			Acquire DB2			Goes to ERROR
Post-condition / Resource group states		WAS: DB2: LDAP:	WAS:ERROR DB2: ERROR LDAP: OFFLINE	WAS: DB2: LDAP:	WAS: DB2: LDAP:	

The home node for LDAP is Node 1. Since the resource group startup policy is Online On Home Node, LDAP could *not* be acquired on Node 2 at node join. Since LDAP is the parent group and is *not* in the ONLINE state, DB2 and WAS are put to ERROR state when the node joins.

Use Case 2: Start First Node (Node 1)

Precondition: As in the post condition of the previous use case.						
Step/ Qualifier	**Action**	**Node 1**	**Node 2**	**Node 3**	**Node 4**	**Comments**
1	Start Node 1					
2		Acquire LDAP				
		Acquire DB2	Acquire DB2			
			Acquire WAS			
Post condition/ Resource group states		WAS: OFFLINE WAS: ONLINE DB2: **ONLINE** LDAP: **ONLINE**	DB2: **ONLINE** LDAP: OFFLINE	WAS: DB2: LDAP:	WAS: DB2: LDAP:	

When Node 1 starts, LDAP is started on Node 1 and Node 2 starts WAS and DB2 since the dependency condition is met.

Use Case 3: Start First Node (Node 3)

Precondition: As in the post condition of the previous use case.						
Step/ Qualifier	**Action**	**Node 1**	**Node 2**	**Node 3**	**Node 4**	**Comments**
1	Start node 3					
2				Acquire WAS_SEC DB2_SEC LDAP_SEC		
Post condition/ Resource group states		WAS: OFFLINE WAS: **ONLINE** DB2: **ONLINE** LDAP: **ONLINE**	DB2: **ONLINE** LDAP: OFFLINE	WAS: **Secondary** DB2: **Secondary** LDAP: **Secondary**	WAS: DB2: LDAP:	

Starting a node in the secondary instance activates all the secondary instances of the resource groups. Starting Node 4 does *not* affect the states of any resource groups. However, Node 4 acquires DB2 in ONLINE_SECONDARY state.

Use Case 4: Fallover of DB2 on Node 1

Precondition: All nodes in the cluster are ONLINE and the state of the resource groups is as follows:						
		Node 1	**Node 2**	**Node 3**	**Node 4**	
		WAS: OFFLINE DB2: **ONLINE** LDAP: **ONLINE**	WAS: **ONLINE** DB2: **ONLINE** LDAP: OFFLINE	WAS: **Secondary** DB2: **Secondary** LDAP: **Secondary**	WAS: OFFLINE DB2: **Secondary** LDAP: OFFLINE	
Step/ Qualifier	**Action**	**Node 1**	**Node 2**	**Node 3**	**Node 4**	**Comments**
1	Disk adapter Failure on Node 1					
2		Release WAS				
		Release DB2	Release DB2			
		Release LDAP				
			Acquire LDAP			
			Acquire DB2			
			Acquire WAS			
Post-condition/ Resource group states		WAS: OFFLINE DB2: OFFLINE LDAP: OFFLINE	WAS: **ONLINE** DB2: **ONLINE** LDAP: **ONLINE**	WAS: **Secondary** DB2: **Secondary** LDAP: **Secondary**	WAS: OFFLINE DB2: **Secondary** LDAP: OFFLINE	

Since LDAP and DB2 should be online on the same node, although DB2 is a concurrent resource group, a failure on Node 1 results in moving LDAP (Online On The Same Node Dependent) to Node 2. Because LDAP is a parent the result is all the resource groups are bounced as a result of adapter failure.

Appendix C: HACMP for AIX Commands

This appendix provides a quick reference to commands commonly used to obtain information about the cluster environment or to execute a specific function. The chapter lists syntax diagrams and provides examples for using each command.

Overview of Contents

As system administrator, you often must obtain information about your cluster to determine if it is operating correctly. The commands you need to obtain this information are listed in alphabetical order in this chapter.

Highlighting

The following highlighting conventions are used in this appendix:

Bold	Identifies command words, keywords, files, directories, and other items whose actual names are predefined by the system.
Italics	Identifies parameters whose actual names or values are supplied by the user.
`Monospace`	Identifies examples of specific data values, examples of text similar to what you might see displayed, examples of program code similar to what you might write as a programmer, messages from the system, or information you should actually type.

Reading Syntax Diagrams

Usually, a command follows this syntax:

[]	Material within brackets is optional.
{}	Material within braces is required.
|	Indicates an alternative. Only one of the options can be chosen.
...	Indicates that one or more of the kinds of parameters or objects preceding the ellipsis can be entered.

Note: Flags listed in syntax diagrams throughout this appendix are those recommended for use with the HACMP for AIX software. Flags used internally by SMIT are *not* listed.

Related Information

For complete information on a command's capabilities and restrictions, see the online man page. Man pages for HACMP for AIX 5L commands and utilities are installed in the **/usr/share/man/cat1** directory. Use the following syntax to read man page information:

```
man [command-name]
```

where *command-name* is the actual name of the HACMP command or script. For example, type **man clpasswd** to obtain information about the HACMP user password command.

Data Collection Utilities

Use the AIX 5L **snap** command to collect data from HACMP clusters.

The **-e** flag collects the hacmp data. Using **/usr/sbin/snap -e** lets you properly document a problem with one simple command. This command gathers all files necessary for determining most HACMP problems. It provides output in a format ready to send to IBM Support personnel.

When you run the command, the output is placed in a newly created subdirectory called **/ibmsupt** in **/tmp**.The recommended amount for available space in /tmp should be at least 100 MG before running the command, the output could require more depending on number of nodes and actual files collected.

See the AIX 5L man page for complete information.

RSCT Command for Testing Disk Heartbeat

The following RSCT command is useful for testing the link status of the disk heartbeating path. The pathname for this command is **/usr/sbin/rsct/bin/dhb_read**.

Note: The disk heartbeating network cannot be tested with this utility while the network is active.

Run the command on both nodes. On one node, run the command with the **-r** flag. This indicates it is in receive mode and will wait for a sign from the remote node. It will *not* wait indefinitely in this mode. On the other node, the utility must be run with the **-t** flag. If the remote node is still in receive mode, and the disk is set up correctly, you should immediately see a message on both nodes that says: Link operating normally.

```
dhb_read -p devicename

dhb_read -p devicename -r

dhb_read -p devicename -t
```

Example Disk Heartbeating Test

To test the disk heartbeating link on nodes A and B, where *hdisk1* is the heartbeat path:

On Node A, enter: `dhb_read -p hdisk1 -r`

On Node B, enter: `dhb_read -p hdisk1 -t`

If the link is active, you see this message on both nodes:

```
Link operating normally.
```

The return code is 0 for success and -1 otherwise.

HACMP for AIX Common Commands

The following commands can be run from the command line to obtain information about your HACMP for AIX cluster environment. Syntax and descriptions of these commands are included in this appendix.

cl_convert	Converts Configuration Database of previous HACMP release to Configuration Database of current release. Run from the command line only if installation fails. Otherwise, it runs automatically with installation.
clconvert_snapshot	Upgrades cluster snapshots.
clRGinfo	Displays the status and location of a given resource group in the cluster configuration.
clgetaddr	Returns IP address for the specified node name.
clpasswd	Changes a user's password on each node in the cluster.
cllscf	Lists cluster topology information.
cllsdisk	Lists PVIDs of accessible disks in a specified resource chain.
cllsfs	List filesystems accessible in a specified resource chain.
cllslv	List the names of filesystems accessible by nodes in a specified resource chain.
cllsgrp	Lists all resource groups.
cllsnim	Lists contents of HACMPnetwork interface module Configuration Database class.
cllsparam	Lists runtime parameters.
cllsres	Lists Configuration Database resource data by name and arguments.
cllsserv	Lists application servers by name.
cllsvg	List volume groups accessible in a specified resource chain.
clshowres	Shows node environment resources.
clstat	Monitors status of cluster.
cltopinfo	Lists all topology information: cluster, nodes, networks, interfaces.
get_local_nodename	Retrieves the name of the local node.
clgetactivenodes	Retrieves the names of all active cluster nodes.
clresactive	Retrieves the names of all active resources.

HACMP for AIX C-SPOC Common Commands

The following C-SPOC commands function in cluster environments and can be run from the command line to manage the cluster. For complete syntax and descriptions of these commands, see HACMP for AIX C-SPOC Commands later in this appendix.

cl_clstop	Stops cluster services on nodes running C-SPOC.
cl_lsfs	Displays shared filesystem attributes for all cluster nodes.
cl_lsgroup	Displays group attributes for all cluster nodes.
cl_lslv	Displays shared logical volume attributes for cluster nodes.
cl_lsuser	Displays user account attributes for all nodes.
cl_lsvg	Displays shared volume group attributes for cluster nodes.
cl_nodecmd	Runs a given command in parallel on a given set of nodes.
cl_rc.cluster	Sets the operating environment and starts cluster daemons.

HACMP for AIX Commands

cl_convert [-F] -v <*release*> [-s<*simulation file*>] [-i]

Upgrading HACMP software to the newest version involves converting the Configuration Database from a previous release to that of the current release. When you install HACMP, **cl_convert** is run automatically. However, if installation fails, you must run **cl_convert** from the command line. Root user privilege is required to run **cl_convert**.

The command copies the previous version's ODM data to the new version's ODM structure. If fields were deleted in the new version, the data is saved to **/tmp/cl_convert_HACMP_OLD**. The command then ensures that the data is in the correct form for the new version.

When the new version is installed, the install script adds the suffix OLD to the HACMPxxx classes stored in the **/etc/objrepos** directory, and it creates the new HACMPxxx classes for the new version. The install script issues the **cl_convert** command which converts the data in HACMPxxxOLD to the corresponding new classes in HACMPxxx.

You may run the **cl_convert** command from the command line, but it is expecting the HACMPxxx and HACMPxxxOLD ODM's to already exist.

You may want to run the **cl_convert** command with the **-F** option. If the option is *not* specified, the **cl_convert** command checks for configured data in the new ODM class HACMPcluster. If data is present, the command exits without performing the conversion. If the **-F** option is specified, the command will continue without checking for present data.

Note that **cl_convert** copies the HACMPxxx and HACMPxxxOLD ODM's to a temporary file (**/tmp/tmpodmdir**) for processing before writing the final data to the HACMPxxx ODM's. If **cl_convert** encounters any kind of error, the HACMPxxx ODM's are not overwritten. If no error occurs, the HACMPxxx ODM's are overwritten and the install script will remove the HACMPxxxOLD ODM's

Note that you must be in the conversion directory to run this command:

/usr/es/sbin/cluster/conversion

Also, `cl_convert` assumes that the correct value for $ODMDIR is set. The results of `cl_convert` can be found in **/tmp/clconvert.log**.

-F	Force flag. Causes `cl_convert` to overwrite existing ODM object classes, regardless of the number of existing entries. Omitting this flag causes `cl_convert` to check for data in HACMPcluster (which there will always be from the previous configuration) and exit if data is encountered.
-v	Release version flag. Indicates the release number of the old version. **WARNING**: Do not use the cl_convert command unless you know the version from which you are converting.
-s <simulation_file>	Simulation flag. Indicates that instead of writing the resulting ODM data back to the new HACMPxxx ODM's, write to the specified file in text format.
-i	Ignore copy flag. Specifies not to copy the `HACMPxxxOLD` data to the new `HACMPxxx` ODM's, but just operate directly on the new `HACMPxxx` ODM's. This is used primarily by `clconvert_snapshot`.

Note: The AIX 5L environmental variable ODMDIR must be set to the directory you wish to convert.

Examples

Example

If a cluster is already configured for HACMP 5.1, during the installation of HACMP 5.4, the installing script will call `cl_convert` as:

```
cl_convert -F -v 5.1
```

clconvert_snapshot -v *release* -s <*snapshot file*>

You can run **clconvert_snapshot** to upgrade cluster snapshots from a previous version (starting from version 5.1) of HACMP to the most recent version of HACMP. The command by default assumes you are converting to the latest version of the software.

The command copies the previous version's ODM data from the snapshot_file to the format of the new version's ODM structure. If fields were deleted in the new version, the data is saved to `/tmp/cl_convert_HACMP_OLD`. The command then ensures that the data is in the correct form for the new version.

Once a snapshot file has been upgraded, it is assigned the same name as the previous version and cannot be reverted back to the previous version. A copy of the old version of the snapshot will be saved for you with the same original name plus the `.old` extension.

You must be in the **/usr/es/sbin/cluster/conversion** directory on the same node that took the snapshot to run the `clconvert_snapshot` command.

Once the snapshot file has been upgraded and all of the nodes in the cluster have the current level installed, the upgraded snapshot can be applied and then the cluster can be brought up.

The script `clconvert_snapshot` creates an old version of the ODMs and populates those ODMs with the values from the user-supplied snapshot file. It then calls the same commands that `cl_convert` uses to convert those ODMs to the current version. A new snapshot is taken from the upgraded ODMs and copied to the user supplied snapshot file.

The **clconvert_snapshot** is *not* run automatically during installation, and must always be run from the command line.

-v	Release version flag. Specifies the release number from which the conversion is to be performed. **WARNING**: Do not use the clconvert_snapshot command unless you know the version from which you are converting.
-s	Snapshot file flag. Specifies the snapshot file to convert. If you do not specify a path, for the snapshot file, the command uses the path specified in the **$SNAPSHOTPATH** variable. The default is **/usr/es/sbin/cluster/snapshots**.

Example

```
clconvert_snapshot -v 5.1 -s mysnapshot
```

Converts an HACMP 5.1 snapshot to an HACMP 5.4 snapshot named “mysnapshot.”

The file “mysnapshot” is in turn placed in the directory specified by the **$SNAPSHOTPATH** environment variable. If a **$SNAPSHOTPATH** variable is *not* specified, the file is put in **/usr/es/sbin/cluster/snapshots**.

clfindres [-s] [resgroup1] [resgroup2]...

Finds a given resource group or groups in a cluster configuration.

Note: When you run **clfindres**, it calls **clRGinfo,** and the command output for **clfindres** is the same as it is for the **clRGinfo** command. Therefore, use the **clRGinfo** command to find the status and the location of the resource groups. See the following section or the man page for the **clRGinfo** command for more information.

-s	Requests abbreviated (location only) output.

clpasswd [-g resource group] *user*

Change the current users password on all nodes in a cluster, or in a resource group.

The Cluster Password (**clpasswd**) utility lets users to change their own password on all nodes in a cluster, or in a resource group as specified by the HACMP administrator, from a single node. Before users can change their password across cluster nodes, the HACMP administrator adds any users who do *not* have root privileges to the list of users allowed to change their password.

For information about giving users permission to change their own password, see the section Allowing Users to Change Their Own Passwords in Chapter 16: Managing User and Groups.

This Cluster Password utility can also replace the AIX 5L password utility from the SMIT fastpath **cl_passwd**.

The following table shows where a user's password is changed based on the user's authorization and the password utility that is active:

	When the system password utility is linked to clpasswd and /bin/passwd is invoked	When the system password utility is active
User authorized to change password across cluster	The password is changed on all cluster nodes,	The password is changed on all cluster nodes.
***User not* authorized to change password across cluster**	The password is changed only on the local node.	The password is *not* changed.

-g	Specifies the name of the resource group in which the user can change their password. The password is changed on each node in the specified resource group.
user	The username of the user who is changing their password.

Example

```
clpasswd -g rg1 myusername
```

clRGinfo [-a][-h] [-v][-s|-c] [-p] [-t] [-d][resgroup1] [resgroup2]...

See the section Using the clRGinfo Command in Chapter 10: Monitoring an HACMP Clusterfor usage and examples.

clgetaddr [-o *odmdir*] *nodename*

Returns a PINGable address for the specified node name.

-o	Specifies an alternate ODM directory.

Example

To get a PINGable address for the node *seaweed*, enter:

```
clgetaddr seaweed
```

The following address is returned: `2361059035`

cllscf

Lists complete cluster topology information. See **cltopinfo** for updated command.

cllsdisk {-g *Resource Group*}

Lists PVIDs of accessible disks in a specified resource chain.

`-g resource group`

Specifies name of resource group to list.

Example

```
cllsdisk -g grp3
```

Lists PVIDs of disks accessible in resource group *grp3*.

cllsfs {-g *resource group*} [-n]

Lists shared filesystems contained in a resource group.

-g *resource group*	Specifies name of resource group for which to list filesystems.
-n	Lists the nodes that share the filesystem in the resource group.

Note: Do *not* run the **cllsfs** command from the command line. Use the SMIT interface to retrieve filesystem information, as explained in Chapter 11: Managing Shared LVM Components.

cllslv [-g *resource group*] [-n] [-v]

Lists the names of logical volumes accessible by nodes in a specified resource chain.

-g *resource group*	Specifies name of resource group for which to list logical volumes.
-n	Lists the nodes in the resource group.
-v	Lists only those logical volumes that belong to volume groups that are currently varied on.

Examples

Example 1

```
cllslv -g grp2
```

Lists all shared logical volumes contained in resource group *grp2*.

Example 2

```
cllslv -g grp2 -n -v
```

Displays the nodes and those logical volumes that belong to currently varied-on volume groups in resource group *grp2*.

cllsgrp

Lists names of all resource groups configured in the cluster.

cllsnim [-d *odmdir*] [-c] [-n *nimname*]

Lists contents of HACMPnim Configuration Database class.

-d *odmdir*	Specifies an alternate ODM directory to **/etc/objrepos**.
-c	Specifies a colon output format.
-n *nimname*	Name of the network interface module for which to list information.

Examples

Example 1
```
cllsnim
```
Shows information for all configured network modules.

Example 2
```
cllsnim -n ether
```
Shows information for all configured Ethernet network modules.

cllsparam {-n *nodename*} [-c] [-s] [-d *odmdir*]

Lists runtime parameters.

-n *nodename*	Specifies a node for which to list the information.
-c	Specifies a colon output format.
-s	Used along with the **-c** flag, specifies native language instead of English.
-d *odmdir*	Specifies an alternate ODM directory.

Example

```
cllsparam -n abalone
```
Shows runtime parameters for node *abalone*.

cllsres [-g *group*] [-c] [-s] [-d *odmdir*] [-q *query*]

Sorts HACMP for AIX Configuration Database resource data by name and arguments.

-g *group*	Specifies name of resource group to list.
-c	Specifies a colon output format.
-s	Used with the **-c** flag, specifies native language instead of English.
-d *odmdir*	Specifies an alternate ODM directory.

-q *query*	Specifies search criteria for ODM retrieve. See the **odmget** man page for information on search criteria.

Examples

Example 1
```
cllsres
```

Lists resource data for all resource groups.

Example 2
```
cllsres -g grp1
```

Lists resource data for resource group *grp1*.

Example 3
```
cllsres -g grp1 -q"name = FILESYSTEM"
```

Lists filesystem resource data for resource group *grp1*.

cllsserv [-c] [-h] [-n *name*] [-d *odmdir*]

Lists application servers by name.

-c	Specifies a colon output format.
-h	Specifies to print a header.
-n *name*	Specifies an application server for which to check information.
-d *odmdir*	Specifies an alternate ODM directory.

Examples

Example 1
```
cllsserv
```

Lists all application servers.

Example 2
```
cllsres -c -n test1
```

Lists information in colon format for application server *test1*.

cllsvg {-g *resource group*} [-n] [-v] [-s | -c]]

Lists volume groups shared by nodes in a cluster. A volume group is considered shared if it is accessible by all participating nodes in a configured resource group. Note that the volume groups listed may or may *not* be configured as a resource in any resource group. If neither **-s** nor **-c** is selected, then both shared and concurrent volume groups are listed.

-g *resource group*	Specifies name of resource group for which to list volume groups that are shared amongst nodes participating in that resource group.
-n *nodes*	Specifies all nodes participating in each resource group.

-v	Lists only volume groups that are varied on, and match other command line criteria.
-s	Lists only shared volume groups that also match other criteria.
-c	Lists only concurrent volume groups that also match other criteria.

Example

```
cllsvg -g grp1
```

Lists all shared volume groups in resource group *grp1*.

clshowres [-g *group*] [-n *nodename*] [-d *odmdir*]

Shows resource group information for a cluster or a node.

-g *group*	Name of resource group to show.
-n *nodename*	Searches the resources Configuration Database from the specified node.
-d *odmdir*	Specifies *odmdir* as the ODM object repository directory instead of the default **/etc/objrepos**.

Examples

Example 1

```
clshowres
```

Lists all the resource group information for the cluster.

Example 2

```
clshowres -n clam
```

Lists the resource group information for node *clam*.

clstat [-c cluster ID | -n cluster name] [-i] [-r seconds] [-a] [-o] [-s]

Cluster Status Monitor (ASCII mode).

-c *cluster id*	Displays cluster information only about the cluster with the specified ID. If the specified cluster is *not* available, **clstat** continues looking for the cluster until the cluster is found or the program is canceled. May *not* be specified if the **-i** option is used.
-i	Runs ASCII **clstat** in interactive mode. Initially displays a list of all clusters accessible to the system. The user must select the cluster for which to display the detailed information. A number of functions are available from the detailed display.
-n *name*	Displays cluster information about the cluster with the specified name. May *not* be specified if the **-i** option is used.

-r *seconds*	Updates the cluster status display at the specified number of seconds. The default is 1 second; however, the display is updated only if the cluster state changes.
-a	Causes **clstat** to display in ASCII mode.
-o	(once) Provides a single snapshot of the cluster state and exits. This flag can be used to run **clstat** out of a **cron** job. Must be run with **-a**; ignores **-i** and **-r** options.
-s	Displays service labels and their state (up or down).

clstat [-a] [-c *id* | *-n name*] [-r *tenths-of-seconds*] [-s]

Cluster Status Monitor (X Windows mode).

-a	Runs **clstat** in ASCII mode.
-c *id*	Displays cluster information only about the cluster with the specified ID. If the specified cluster is *not* available, **clstat** continues looking for the cluster until the cluster is found or the program is canceled. May *not* be specified if the **-n** option is used.
-n *name*	Displays cluster information only about the cluster with the specified name.
-r *tenths-of-seconds*	The interval at which the **clstat** utility updates the display. For the graphical interface, this value is interpreted in tenths of seconds. By default, **clstat** updates the display every 0.10 seconds.
-s	Displays service labels and their state (up or down).

Examples

Example 1

```
clstat -n mycluster
```

Displays the cluster information about the cluster named *mycluster*.

Example 2

```
clstat -i
```

Runs ASCII **clstat** in interactive mode, allowing multi-cluster monitoring.

Buttons on X Window System Display

Prev	Displays previous cluster.
Next	Displays next cluster.
Name:Id	Refresh bar, pressing bar causes **clstat** to refresh immediately.
Quit	Exits application.
Help	Pop-up help window shows the **clstat** manual page.

cltopinfo [-c] [-i] [-n] [-w]

Shows complete topology information: The cluster name, total number of networks and nodes configured in the cluster. Displays all the configured networks for each node. Displays all the configured interfaces for each network. Also displays all the resource groups defined.

-c	Shows the cluster name and the security mode (Standard or Enhanced)
-i	Shows all interfaces configured in the cluster. The information includes the interface label, the network it's attached to (if appropriate), the IP address, netmask, nodename and the device name.
-n	Shows all the nodes configured in the cluster. For each node, lists all the networks defined. For each network, lists all the interfaces defined and the distribution preference for service IP label aliases (if defined)—this is new.
-w	Shows all the networks configured in the cluster. For each network, lists all the nodes attached to that network. For each node, lists all the interfaces defined and the distribution preference for service IP label aliases (if defined)—this is new.

Examples

Example 1

To show all the nodes and networks defined in the cluster (nodes abby and polly):

```
# cltopinfo
        Cluster Description of Cluster c10
        Cluster Security Level Standard
        There are 2 node(s) and 4 network(s) defined

NODE polly:
                Network net_ether_01
                        polly_en1stby    192.168.121.7
                        polly_en0boot    192.168.120.7
                Network net_ether_02
               Network net_ether_02 will collocate service label(s) with
the persistent label (if any).

                Network net_rs232_01
                Network net_rs232_02
                      polly_tty0_01    /dev/tty0
```

```
NODE abby:
        Network net_ether_01
               abby_en0boot    192.168.120.9
               abby_en1stby    192.168.121.9
               abby_en2boot    192.168.122.9
        Network net_ether_02
        Network net_rs232_01
        Network net_rs232_02
               abby_tty0_01    /dev/tty0
No resource groups defined
```

Example 2

To show the cluster name and current security mode:

```
# cltopinfo -c

    Cluster Description of Cluster c10
    Cluster Security Level Standard.
```

Example 3
To show all the interfaces defined in the cluster (nodes are nip and tuck):

```
# cltopinfo -i

  If Name          Network         Type    Node    Address          If   Netmask
  =========       ============    ====    ====    =============    ===  =============
  nip_en1stby     net_ether_01    ether   nip     192.168.121.7    en2  255.255.255.0
  nip_en0boot     net_ether_01    ether   nip     192.168.120.7    en1  255.255.255.0
  nip_tty0_01     net_rs232_02    rs232   nip     /dev/tty0        tty0
  tuck_en0boot    net_ether_01    ether   tuck    192.168.120.9    en1  255.255.255.0
  tuck_en1stby    net_ether_01    ether   tuck    192.168.121.9    en2  255.255.255.0
  tuck_en2boot    net_ether_01    ether   tuck    192.168.122.9    en3  255.255.255.0
  tuck_tty0_01    net_rs232_02    rs232   tuck    /dev/tty0        tty0
```

get_local_nodename

Returns the name of the local node.

clgetactivenodes [-n *nodename*] [-o *odmdir*] [-t *timeout*] [-v *verbose*]

Retrieves the names of all cluster nodes.

-n *nodename*	Determines if the specified node is active.
-o *odmdir*	Specifies ***odmdir*** as the ODM object repository directory instead of the default **/etc/objrepos**.
-t *timeout*	Specifies a maximum time interval for receiving information about active nodes.
-v *verbose*	Specifies that information about active nodes be displayed as verbose output.

Example

```
clgetactivenodes -n java
```

Verifies that node *java* is active.

clresactive {-v *volumegroup* | -l *logicalvolume* | - f *filesystem* |-u *user* |-g *group* |-V *HACMP version* |-c [*:cmd*]}

Retrieves the names of all active cluster resources.

-v *volumegroup*	Specifies the status of a volume group.
-l *logicalvolume*	Specifies the status of a logical volume.
-f *filesystem*	Specifies the status of a filesystem.
-u *user*	Specifies a user account.
-g *group*	Specifies a user group.
-V *HACMP version*	Specifies the current HACMP for AIX version.

-c*::cmd*	Specifies several commands to be executed simultaneously.

Example

```
clresactive -g  finance
```

HACMP for AIX C-SPOC Commands

The following C-SPOC commands can be executed from the command line and through SMIT. Error messages and warnings returned by the commands are based on the underlying AIX-related commands.

Note: While the AIX 5L commands, underlying the C-SPOC commands, allow you to specify flags in any order, even flags that require arguments, the C-SPOC commands require that arguments to command flags must immediately follow the flag. See the **cl_lsuser** command for an example.

```
cl_clstop [-cspoc "[-f] [-g ResourceGroup | -n NodeList] "] -f
cl_clstop [-cspoc "[-f] [-g ResourceGroup | -n NodeList] "] -g [-s] [-y]
[-N | -R | -B]
cl_clstop [-cspoc "[-f] [-g ResourceGroup | -n NodeList] "] -gr [-s]
[-y] [-N | -R |-B]
```

Stops Cluster daemons using the System Resource Controller (SRC) facility.

Note: Arguments associated with a particular flag must be specified immediately following the flag.

-cspoc	Argument used to specify one of the following C-SPOC options: **-f** – Forces **cl_stop** to skip default verification. If this flag is set and a cluster node is *not* accessible, **cl_clstop** reports a warning and continues execution on the other nodes. **-g *ResourceGroup*** – Generates the list of nodes participating in the resource group where the command will be executed. **-n *NodeList*** – Shuts down cluster services on the nodes specified in the nodelist.
-f	Cluster services are stopped and resource groups being placed in and UNMANAGED state. Cluster daemons should terminate without running any local events. Resources are *not* released.
-g	Cluster services are stopped with resource groups brought offline. Resources are *not* released.
-gr	Cluster services are stopped with resource groups moved to another node, if configured. The daemons should terminate gracefully, and the node should release its resources, which will then be taken over. A nodelist must be specified for cluster services to be stopped with resource groups moved to another node.

-s	Silent shutdown, specifies *not* to broadcast a shutdown message through **/bin/wall**. The default is to broadcast.
-y	Do *not* ask operator for confirmation before shutting down.
-N	Shut down now.
-R	Stops on subsequent system restart (removes the **inittab** entry).
-B	Stop now.

Examples

Example 1
To shut down the cluster services with resource groups brought offline on *node1* (releasing the resources) with no warning broadcast to users before the cluster processes are stopped and resources are released, enter:

```
cl_clstop -cspoc "-n node1" -gr -s -y
```

Example 2
To shut down the cluster services and place resource groups in an UNMANAGED state on all cluster nodes (resources *not* released) with warning broadcast to users before the cluster processes are stopped, enter:

```
cl_clstop -f -y
```

Example 3
To shut down the cluster services with resource groups brought offline on all cluster nodes, broadcasting a warning to users before the cluster processes are stopped, enter:

```
cl_clstop -g -y
```

cl_lsfs [-cspoc"[-f] [-g *ResourceGroup* | -n *Nodelist*]" [-q] [-c | -l] *FileSystem*]...

Displays the characteristics of shared filesystems.

Note: Arguments associated with a particular flag must be specified immediately following the flag.

-cspoc	Argument used to specify one of the following C-SPOC options: **-f** – This option has no effect when used with the **cl_lsfs** command. **-g *ResourceGroup*** – Generates the list of nodes participating in the resource group where the command will be executed. **-n *nodelist*** – Runs the command on this list of nodes. If more than one node, separate nodes listed by commas.
-c	Specifies a different search pattern to determine if the underlying AIX 5L **lsfs** command returned data or *not*.
-l	Specifies that the output should be in list format.

-q	Queries the logical volume manager (LVM) for the logical volume size (in 512-byte blocks) and queries the JFS superblock for the filesystem size, the fragment size, the compression algorithm (if any), and the number of bytes per i-node (nbpi). This information is displayed in addition to other filesystem characteristics reported by the **lsfs** command.

Example

Example 1
To display characteristics about all shared filesystems in the cluster, enter:

```
cl_lsfs
```

Example 2
Display characteristics about the filesystems shared amongst the participating nodes in *resource_grp1*.

```
cl_lsfs -cspoc "-g resource_grp1"
```

cl_lsgroup [-cspoc "[-f] -g *ResourceGroup* | -n *Nodelist*"] [-c|-f] [-a | -a *List*] {ALL | Group [,*Group] ...*}

Displays attributes of groups that exist on an HACMP cluster.

Note: Arguments associated with a particular flag must be specified immediately following the flag.

-cspoc	Argument used to specify the following C-SPOC option: **-f**—This option has no effect when used with the **cl_lsgroup** command. **-g *ResourceGroup***—Generates the list of nodes participating in the resource group where the command will be executed. **-n *nodelist***—Runs the command on this list of nodes. If more than one node, separate nodes listed by commas.
-a *List*	Specifies the attributes to display. The ***List*** parameter can include any attribute defined in the **chgroup** command, and requires a blank space between attributes. If you specify an empty list using only the **-a** flag, only the group names are listed.

-c	Displays the attributes for each group in colon-separated records, as follows:
	`# name: attribute1: attribute2:...`
	`Group: value1:  value2:  ...`

-f Displays the group attributes in stanzas. Each stanza is identified by a group name. Each Attribute=Value pair is listed on a separate line:

```
group:
attribute1=value
attribute2=value
attribute3=value
```

ALL | group [group]... All resource groups, or particular group or groups to display.

Examples

Example 1
To display the attributes of the finance group from all cluster nodes enter:

```
cl_lsgroup finance
```

Example 2
To display in stanza format the ID, members (users), and administrators (adms) of the finance group from all cluster nodes, enter:

```
cl_lsgroup -f -a id users adms finance
```

Example 3
To display the attributes of all the groups from all the cluster nodes in colon-separated format, enter:

```
cl_lsgroup -c ALL
```

All the attribute information appears, with each attribute separated by a blank space.

cl_lslv [-cspoc "[-f] [-g *ResourceGroup* | -n *Nodelist*"]] [-l | -m] *LogicalVolume*

Displays shared logical volume attributes.

Note: Arguments associated with a particular flag must be specified immediately following the flag.

-cspoc Argument used to specify one of the following C-SPOC options:

-f – This option has no effect when used with the **cl_lsfs** command.

-g *ResourceGroup* – Generates the list of nodes participating in the resource group where the command will be executed.

-n *Nodelist* – Runs the command on this list of nodes. If more than one node, separate nodes listed by commas.

-l *Logical Volume* Lists information for each physical volume in the shared logical volume. Refer to the **lslv** command for information about the fields displayed.

-m *Logical Volume* Lists information for each logical partition. Refer to the **lslv** command for information about the fields displayed. If no flags are specified, information about the shared logical volume and its underlying shared volume group is displayed. Refer to the **lslv** command for the information about the fields displayed.

Examples

Example 1
To display information about the shared logical volume *lv03*, enter:

```
cl_lslv -cspoc -g resource_grp1 lv03
```

Information about logical volume *lv03*, its logical and physical partitions, and the volume group to which it belongs is displayed.

Example 2
To display information about a specific logical volume, using the identifier, enter:

```
cl_lslv -g resource_grp1 00000256a81634bc.2
```

All available characteristics and status of this logical volume are displayed.

cl_lsuser [-cspoc "[-f] [-g *ResourceGroup* | -n *Nodelist*]"] [-c | -f] [-a *List*] {ALL | *Name* [,*Name*]...}

Displays user account attributes for users that exist on an HACMP cluster.

Note: Arguments associated with a particular flag must be specified immediately following the flag.

-cspoc Argument used to specify the following C-SPOC option:

-f – This option has no effect when used with the **cl_lsuser** command.

-g *ResourceGroup* – Generates the list of nodes participating in the resource group where the command will be executed.

-n *Nodelist* – Runs the command on this list of nodes. If more than one node, separate nodes listed by commas.

-a *Lists* Specifies the attributes to display. The List variable can include any attribute defined in the **chuser** command and requires a blank space between attributes. If you specify an empty list, only the user names are displayed.

-c Displays the user attributes in colon-separated records, as follows:

```
# name:  attribute1:  attribute2:  ...
User:      value1:      value2:      ...
```

-f Displays the output in stanzas, with each stanza identified by a user name. Each Attribute=Value pair is listed on a separate line:

```
user:
attribute1=value
attribute2=value
attribute3=value
```

ALL | Name [name]... Display information for all users or specified user or users.

Examples

Example 1
To display in stanza format the user ID and group-related information about the *smith* account from all cluster nodes, enter:

```
cl_lsuser -fa id pgrp groups admgroups smith
```

Example 2
To display all attributes of user *smith* in the default format from all cluster nodes, enter:

```
cl_lsuser smith
```

All attribute information appears, with each attribute separated by a blank space.

Example 3
To display all attributes of all the users on the cluster, enter:

```
cl_lsuser ALL
```

All attribute information appears, with each attribute separated by a blank space.

cl_lsvg [-cspoc "[-f] [-g *ResourceGroup* | n- *Nodelist*]" [-o] | [-l | -M | -p] *Volume Group...*

Displays information about shared volume groups.

Note: Arguments associated with a particular flag must be specified immediately following the flag.

-cspoc Argument used to specify one of the following C-SPOC options:

-f – This option has no effect when used with the **cl_lsvg** command.

-g *ResourceGroup* – Specifies the name of the resource group whose participating nodes share the volume group. The command executes on these nodes.

-n *Nodelist* – Runs the command on this list of nodes. If more than one node, separate nodes listed by commas.

-p	Lists the following information for each physical volume within the group specified by the ***VolumeGroup*** parameter: **- Physical volume:** A physical volume within the group. **- PVstate**: State of the physical volume. - **Total PPs**: Total number of physical partitions on the physical volume. **- Free PPs**: Number of free physical partitions on the physical volume. - **Distribution**: The number of physical partitions allocated within each section of the physical volume: outer edge, outer middle, center, inner middle, and inner edge of the physical volume.
-l	Lists the following information for each logical volume within the group specified by the ***VolumeGroup*** parameter: **- LV**: A logical volume within the volume group. **- Type**: Logical volume type. **- LPs**: Number of logical partitions in the logical volume. **- PPs:** Number of physical partitions used by the logical volume. **- PVs**: Number of physical volumes used by the logical volume.
-M	Lists the following fields for each logical volume on the physical volume: **- PVname: *PPnum* [*LVname*: *LPnum* [:*Copynum*] [*PPstate*]]** - **PVname**: Name of the physical volume as specified by the system. - **PPnum**: Physical partition number. Physical partition numbers can range from 1 to 1016.
-o	Lists only the active volume groups (those that are varied on). An active volume group is one that is available for use. Refer to the **lsvg** command for the information displayed if no flags are specified.

Examples

Example 1
To display the names of all shared volume groups in the cluster, enter:

```
cl_lsvg
nodeA: testvg
nodeB: testvg
```

Example 2
To display the names of all active shared volume groups in the cluster, enter:

```
cl_lsvg -o
nodeA: testvg
```

The **cl_lsvg** command lists only the node on which the volume group is varied on.

Example 3
To display information about the shared volume group *vg02*, enter:

```
cl_lsvg -cspoc testvg
```

The **cl_lsvg** command displays the same data as the **lsvg** command, prefixing each line of output with the name of the node on which the volume group is varied on.

cl_nodecmd [-q] [-cspoc “[-f] [-n *nodelist* | -g *resource group*]”] *command args*

Runs a given command in parallel on a given set of nodes.

-q	Specifies quiet mode. All standard output is suppressed.
-cspoc	Argument used to specify one of the following C-SPOC options: **-f** – Forces **cl_nodecmd** to skip HACMP version compatibility checks and node accessibility verification. **-g *resource group*** – Generates the list of nodes participating in the resource group where the command will be executed. **-n *nodelist*** – Runs the command on this list of nodes. If more than one node, separate nodes listed by commas.
command	Specifies the command to be run on all nodes in the nodelist.
args	Specifies any arguments required for use with the **cl_nodecmd** command.

Examples

Example 1

```
cl_nodecmd lspv
```

Runs the **lspv** command on all cluster nodes.

Example 2

```
cl_nodecmd -cspoc "-n beaver,dam" lsvg rootvg
```

Runs the **lsvg rootvg** command on nodes *beaver* and *dam*, suppressing standard output.

cl_rc.cluster [-cspoc “[-f] [-g *ResourceGroup* | -n *NodeList*]”] [-boot] [b] [-i] [-N | -R | -B]

Sets up the operating system environment and starts the cluster daemons across cluster nodes.

Note: Arguments associated with a particular flag must be specified immediately following the flag.

-cspoc	Argument used to specify the following C-SPOC option: **-f** – Forces **cl_rc.cluster** to skip HACMP version compatibility checks and node accessibility verification. **-g *ResourceGroup*** – Specifies the name of the resource group whose participating nodes share the volume group. The command executes on these nodes. **-n *Nodelist*** – Executes underlying AIX 5L commands across nodes in the nodelist.
-boot	Configures the service network interface to be on its boot address if IPAT is enabled.
-i	Starts the Cluster Information (**clinfoES**) daemon with its default options.
-b	Broadcasts the startup.
-N	Starts the daemons immediately (no **inittab** change).
-R	Starts the HACMP daemons on system restart only (the HACMP startup command is added to **inittab** file).
-B	Starts the daemons immediately and adds the HACMP entry to the **inittab** file.
-f	Forced startup. Cluster daemons should initialize running local procedures.

Examples

Example 1

To start the cluster with **clinfo** running on all the cluster nodes, enter:

```
cl_rc.cluster -boot -i
```

Appendix D: RSCT: Resource Monitoring and Control Subsystem

Starting with HACMP 5.3, HACMP interfaces with the RSCT Resource Monitoring and Control (RMC) subsystem instead of with the Event Management subsystem. In previous versions of the HAES product, the Cluster Manager interfaces with the Event Manager to let you define custom events based on system metrics and an expression that applied to those metrics. It also uses the Event Manager to monitor applications as well as to determine the nodelist for resource groups if dynamic node priority is used.

The RMC subsystem replaces Event Management resource variables with resource classes and attributes associated with the resource class.

A resource class is a set of resources of the same type. A resource class defines the common characteristics that instances of the resource class can have. A resource attribute describes some characteristic of a resource. There are two types of resource attributes, persistent attributes and dynamic attributes. Persistent attributes describe enduring characteristics of the resource. Dynamic attributes, on the other hand, represent changing characteristics of the resource.

See the *RSCT for AIX 5L and Linux: Administration Guide* and the *RSCT for AIX 5L Technical Reference* for more information.

HACMP now uses the RMC subsystem for these instances:

- Dynamic Node Priority
- Application Monitoring
- User Defined Events.

Conversion of User-Defined Event Definitions

The **clconvert** routine only converts a subset of Event Manager user-defined event definitions to the corresponding RMC event definitions. All user-defined event definitions must be manually reconfigured with the exception of the following UDE definitions defined by DB2:

```
HACMPude:
        name = "DB2_PROC_DOWN"
        state = 0
        recovery_prog_path = "/usr/bin/db2_proc_recovery"
        recovery_type = 2
        recovery_level = 0
        res_var_name = "IBM.PSSP.Prog.xpcount"
        instance_vector = "NodeNum=*;ProgName=db2sysc;UserName=root"
        predicate = "X@0<X@1 && X@1!=0"
        rearm_predicate = "X@0>X@1"

HACMPude:
        name = "DB2_PAGINGSPACE_LOW"
        state = 0
        recovery_prog_path = "/usr/bin/db2_paging_space"
        recovery_type = 2
        recovery_level = 0
        res_var_name = "IBM.PSSP.aixos.PagSp.%totalfree"
```

```
        instance_vector = "NodeNum=*"
        predicate = "X<31"
        rearm_predicate = "X>70"

HACMPude:
        name = "DB2_PAGINGSPACE_VERY_LOW"
        state = 0
        recovery_prog_path = "/usr/bin/db2_paging_space"
        recovery_type = 2
        recovery_level = 0
        res_var_name = "IBM.PSSP.aixos.PagSp.%totalfree"
        instance_vector = "NodeNum=*"
        predicate = "X<11"
        rearm_predicate = "X>90"

HACMPude:
        name = "DB2_AUTO_PROC_DOWN"
        state = 0
        recovery_prog_path = "/usr/bin/nfs_auto_recovery"
        recovery_type = 2
        recovery_level = 0
        res_var_name = "IBM.PSSP.Prog.xpcount"
        instance_vector = "NodeNum=*;ProgName=automount;UserName=root"
        predicate = "X@0<X@1 && X@1!=0"
        rearm_predicate = "X@0>X@1"

HACMPude:
        name = "DB2_NFSSTAT_PROC_DOWN"
        state = 0
        recovery_prog_path = "/usr/bin/nfs_auto_recovery"
        recovery_type = 2
        recovery_level = 0
        res_var_name = "IBM.PSSP.Prog.xpcount"
        instance_vector = "NodeNum=*;ProgName=rpc.statd;UserName=root"
        predicate = "X@0<X@1 && X@1!=0"
        rearm_predicate = "X@0>X@1"

HACMPude:
        name = "DB2_NFSLOCK_PROC_DOWN"
        state = 0
        recovery_prog_path = "/usr/bin/nfs_auto_recovery"
        recovery_type = 2
        recovery_level = 0
        res_var_name = "IBM.PSSP.Prog.xpcount"
        instance_vector = "NodeNum=*;ProgName=rpc.lockd;UserName=root"
        predicate = "X@0<X@1 && X@1!=0"
        rearm_predicate = "X@0>X@1"
HACMPude:
        name = "DB2_NFS_PROC_DOWN"
        state = 0
        recovery_prog_path = "/usr/bin/nfs_auto_recovery"
        recovery_type = 2
        recovery_level = 0
        res_var_name = "IBM.PSSP.Prog.xpcount"
        instance_vector = "NodeNum=*;ProgName=nfsd;UserName=root"
        predicate = "X@0<X@1 && X@1!=0"
        rearm_predicate = "X@0>X@1"
```

Conversion of Event Manager-based Dynamic Node Priority Policy and Application Monitor Definitions

The existing preconfigured dynamic node priority policy definitions and the existing Application Monitor definitions are converted to the equivalent RMC definitions. All other user-defined node priority policies are converted to use the default policy (the ordered nodelist). Only the three preconfigured dynamic node priority policies are supported in HACMP 5.3 and 5.4.

Mapping of Event Management Resource Variables to RMC Resource Attributes

The following table shows how the Event Management resource variables map to the RMC resource attributes.

Event Manager Resource Variable	RMC Resource Attribute	Description
IBM.PSSP.aixos.PagSp.totalfree	IBM.Host.TotalPgSpFree	Available virtual memory paging space
IBM.PSSP.aixos.Disk.busy	IBM.PhysicalVolume.PctBusy	Fraction of time disks are busy
IBM.PSSP.aixos.CPU.glidle	IBM.Host. PctTotalTimeIdle	Available processor time
IBM.PSSP.aixos.CPU.glkern	IBM.Host.PctTotalTimeKernel	Amount of time all processors is in kernel mode
IBM.PSSP.aixos.CPU.gluser	IBM.Host.PctTotalTimeUser	Amount of time all processors are in user mode
IBM.PSSP.aixos.CPU.glwait	IBM.Host.PctTotalTimeWait	Amount of time all processors are in wait state
IBM.PSSP.aixos.cpu.idle	IBM.Processor.PctTimeIdle	Amount of time a processor is in idle state
IBM.PSSP.aixos.cpu.kern	IBM.Processor.PctTimeKernel	Amount of time a processor is in kernel mode
IBM.PSSP.aixos.cpu.user	IBM.Processor.PctTimeUser	Amount of time a processor is in user mode
IBM.PSSP.aixos.cpu.wait	IBM.Processor.PctTimeWait	Amount of time a processor is in wait state
IBM.PSSP.aixos.Disk.rblk	IBM.PhysicalVolume.RdBlkRate	Average rate at which blocks are read from the disk
IBM.PSSP.aixos.Disk.wblk	IBM.PhysicalVolume.WrBlkRate	Average rate at which blocks are written to the disk
IBM.PSSP.aixos.Disk.xfer	IBM.PhysicalVolume.XferRate	Average rate of transfers issued to the disk
IBM.PSSP.aixos.FS.%nodesused	IBM.FileSystem.PercentINodeUsed	Percentage of total I-Node used
IBM.PSSP.aixos.FS.%totused	IBM.FileSystem.PercentTotUsed	Percentage of total space used

Event Manager Resource Variable	RMC Resource Attribute	Description
IBM.PSSP.aixos.VG.free	Unavailable	The amount of free space in a volume group, in megabytes.
IBM.PSSP.aixos.LAN.rcverrors	Unavailable	Count of frame receive errors at adapter level.
IBM.PSSP.aixos.LAN.recvdrops	Unavailable	Count of receive packets dropped at device driver level.
IBM.PSSP.aixos.LAN.xmitdrops	Unavailable	Count of transmit packets dropped at device driver level.
IBM.PSSP.aixos.LAN.xmiterrors	Unavailable	Count of frame transmit errors at adapter level.
IBM.PSSP.aixos.LAN.xmitovfl	Unavailable	Count of transmit queue overflows.
IBM.PSSP.aixos.Mem.Kmem.calls	IBM.Host.KMemNumMbuf	Average number of allocated Mbufs
IBM.PSSP.aixos.Mem.Kmem.failures	Unavailable	Number of unsuccessful requests for a kernel memory buffer.
IBM.PSSP.aixos.Mem.Kmem.inuse	Unavailable	Count of kernel memory buffers in use.
IBM.PSSP.aixos.Mem.Kmem.memuse	Unavailable	Current memory use (bytes).
IBM.PSSP.aixos.Mem.Real.%free	Unavailable	Percent memory that is free.
IBM.PSSP.aixos.Mem.Real.%pinned	Unavailable	Percent memory that is pinned.
IBM.PSSP.aixos.Mem.Real.numfrb	Unavailable	Number of pages on free list.
IBM.PSSP.aixos.Mem.Real.size	Unavailable	Size of physical memory (4K pages).
IBM.PSSP.aixos.Mem.Virt.pagein	Unavailable	4K pages read by VMM.
IBM.PSSP.aixos.Mem.Virt.pageout	Unavailable	4K pages written by VMM.
IBM.PSSP.aixos.Mem.Virt.pagexct	Unavailable	Total page faults.
IBM.PSSP.aixos.Mem.Virt.pgspgin	Unavailable	4K pages read from paging space by VMM.
IBM.PSSP.aixos.Mem.Virt.pgspgout	Unavailable	4K pages written to paging space by VMM.
IBM.PSSP.aixos.PagSp.%totalfree	IBM.Host.TotalPgSpFree	Total free disk paging space (4K pages).
IBM.PSSP.aixos.PagSp.%totalused	IBM.Host.PctTotalPgSpUsed	Total used disk paging space (percent).
IBM.PSSP.aixos.PagSp.totalsize	IBM.Host.TotalActivePagingSpace	Total active paging space size (4K pages).

Event Manager Resource Variable	RMC Resource Attribute	Description
IBM.PSSP.aixos.pagsp.%free	IBM.PagingDevice.PctFree	Free portion of this paging space LV (percent).
IBM.PSSP.aixos.pagsp.size	IBM.PagingDevice.Size	Size of a paging space LV (4K pages).
IBM.PSSP.aixos.Proc.runque	IBM.Host.ProcRunQueueb	Average count of processes that are waiting for the CPU
IBM.PSSP.aixos.Proc.swpque	IBM.Host.ProcSwapQueue	The average number of processes waiting to be paged in.
IBM.PSSP.Prog.pcount	IBM.Program.ProgramName	Processes running a specified program.
IBM.PSSP.Prog.xpcount	IBM.Program.ProgramName	Processes running a specified program, called by an executable routine.

Appendix E: Using DLPAR and CUoD in an HACMP Cluster

This appendix describes how to configure and use HACMP in a hardware and software configuration that uses *Dynamic Logical Partitions (DLPARs)* and the *Capacity Upgrade on Demand (CUoD)* function.

The topics in this appendix include:

- Overview and Related Documentation
- LPAR, DLPAR, and CUoD Terminology and Terminology for Resource Types and Memory Allocation
- HACMP Integration with the CUoD Function
- Planning for CUoD and DLPAR and Software and Hardware Requirements for CUoD and DLPAR
- Configuring CUoD in HACMP
- Changing the DLPAR and CUoD Resources Dynamically
- How Application Provisioning Works in HACMP
- Using Pre- and Post-Event Scripts
- Troubleshooting DLPAR and CUoD Operations in HACMP.

Overview

The IBM pSeries servers let you configure multiple *Logical Partitions (LPARs)* on a single physical frame, where each of the LPARs behaves as a standalone pSeries processor. Using this configuration, you can install and run multiple applications on different LPARs that use a single physical hardware component. The applications running on LPARs are completely isolated from each other at the software level. Each LPAR can be optimally tuned for a particular application that runs on it.

In addition, *Dynamic Logical Partitioning (DLPAR)* allows you to dynamically allocate additional resources (such as memory and CPUs) to each logical partition, if needed, without stopping the application. These additional resources must be physically present on the frame that uses logical partitions.

Capacity Upgrade on Demand is one of the features of the DLPAR function that lets you activate preinstalled but yet inactive (and unpaid for) processors as resource requirements change.

Related Documentation

The following related publications provide more information:

- DLPAR information: *Planning for Partitioned-System Operations, SA38-0626*

- CUoD information: *IBM eServer pSeries Planning Guide for Capacity Upgrade on Demand*
- HMC information*: IBM Hardware Management Console for pSeries Installation and Operations Guide, SA38-0590.*

These guides and other related eServer pSeries documentation, are available at:

http://publib16.boulder.ibm.com/pseries/en_US/infocenter/base/

The following whitepapers provide useful information:

- Minimizing Downtime by Using HACMP in a Single Frame Environment,

  http://www.ibm.com/servers/eserver/pseries/software/whitepapers/hacmp_lpar.pdf
- Dynamic Logical Partitioning in IBM eServer pSeries,

  http://www.ibm.com/servers/eserver/pseries/hardware/whitepapers/dlpar.pdf

LPAR, DLPAR, and CUoD Terminology

The appendix uses the terms listed in this section. For more information on the terms, refer to the IBM guides on LPARs and CUoD listed in the Related Documentation section.

- **Logical Partition (LPAR)**. The division of a computer's processors, memory, and hardware resources into multiple environments so that each environment can be operated independently with its own operating system and applications.

  The number of logical partitions that can be created depends on the system. Typically, partitions are used for different purposes, such as database operation, client/server operations, Web server operations, test environments, and production environments. Each partition can communicate with the other partitions as if each partition is a separate machine.
- **Dynamic Logical Partitioning (DLPAR)**. A facility in some pSeries processors that provides the ability to logically attach and detach a *managed system's* resources to and from a logical partition's operating system without rebooting. Some of the features of DLPAR include:
  - **Capacity Upgrade on Demand (CUoD)**, a feature of the pSeries, which allows you to activate preinstalled but inactive processors as resource requirements change.
  - The **Dynamic Processor Deallocation** feature of the pSeries servers, and on some SMP models. It lets a processor to be taken offline dynamically when an internal threshold of recoverable errors is exceeded. DLPAR allows to substitute the inactive processor, if one exists, for the processor that is suspected of being defective. This online switch does *not* impact applications and kernel extensions. This function is *not* supported by HACMP. See the *IBM's Planning for Partitioned-System Operations Guide.*
  - **Cross-partition workload management**, which is particularly important for server consolidation in that it can be used to manage system resources across partitions. This function is *not* supported by HACMP. See the *IBM's Planning for Partitioned-System Operations Guide.*

- **Capacity Upgrade on Demand (CUoD or COD)**. A facility in some pSeries processors that lets you acquire— but *not* pay for—a fully configured system. The additional CPUs and memory, while physically present, are *not* used until you decide that the additional capacity you need is worth the cost. This provides you with a fast and easy upgrade in capacity to meet peak or unexpected loads.
- **Hardware Management Console (HMC)**. An interface that lets you manage all DLPAR operations on several or all LPARs created on the frame, collect CUoD system profile information and enter *activation codes* for CUoD. For integration with HACMP, HMC should have a TCP/IP connection to the LPAR and a configured IP label through which a connection will be established. The **lshmc** command displays the HMC configuration.
- **Managed System**. A pSeries frame that is LPAR-capable and that is managed by an HMC.
- **On/Off Capacity Upgrade on Demand (CUoD)**. A type of CUoD license allowing temporary activation of processors *only*. For more information, see Types of CUoD Licenses.
- **Trial Capacity Upgrade on Demand (CUoD)**. A type of CUoD license allowing a no-charge usage for testing inactive CUoD processors and memory. For more information, see Types of CUoD Licenses.
- **CUoD Vital Product Data (VPD)**. A collection of system profile information that describes the hardware configuration and identification numbers. In this document, use of the term VPD refers to CUoD VPD.
- **Activation Code (or License Key)**. A password used to activate inactive (standby) processors or memory in CUoD. Each activation code is uniquely created for a system and requires the system *VPD (Vital Product Data)* to ensure correctness.

  Note: In HACMP SMIT and in this guide, the activation code is also referred to as the *license key*.

HACMP Integration with the CUoD Function

This section describes how HACMP integrates with the DLPAR and CUoD facilities.

By integrating with DLPAR and CUoD, HACMP ensures that each node can support the application with reasonable performance at a minimum cost. This way, you can upgrade the capacity of the logical partition in cases when your application requires more resources, without having to pay for idle capacity until you actually need it.

You can configure cluster resources so that the logical partition with minimally allocated resources serves as a standby node, and the application resides on another LPAR node that has more resources than the standby node. This way, you do *not* use any additional resources that the frames have until the resources are required by the application.

When it is necessary to run the application on the standby node, HACMP ensures that the node has sufficient resources to successfully run the application. The resources can be allocated from two sources:

- The free pool. The DLPAR function provides the resources to the standby node, by allocating the resources available in the free pool on the frame.

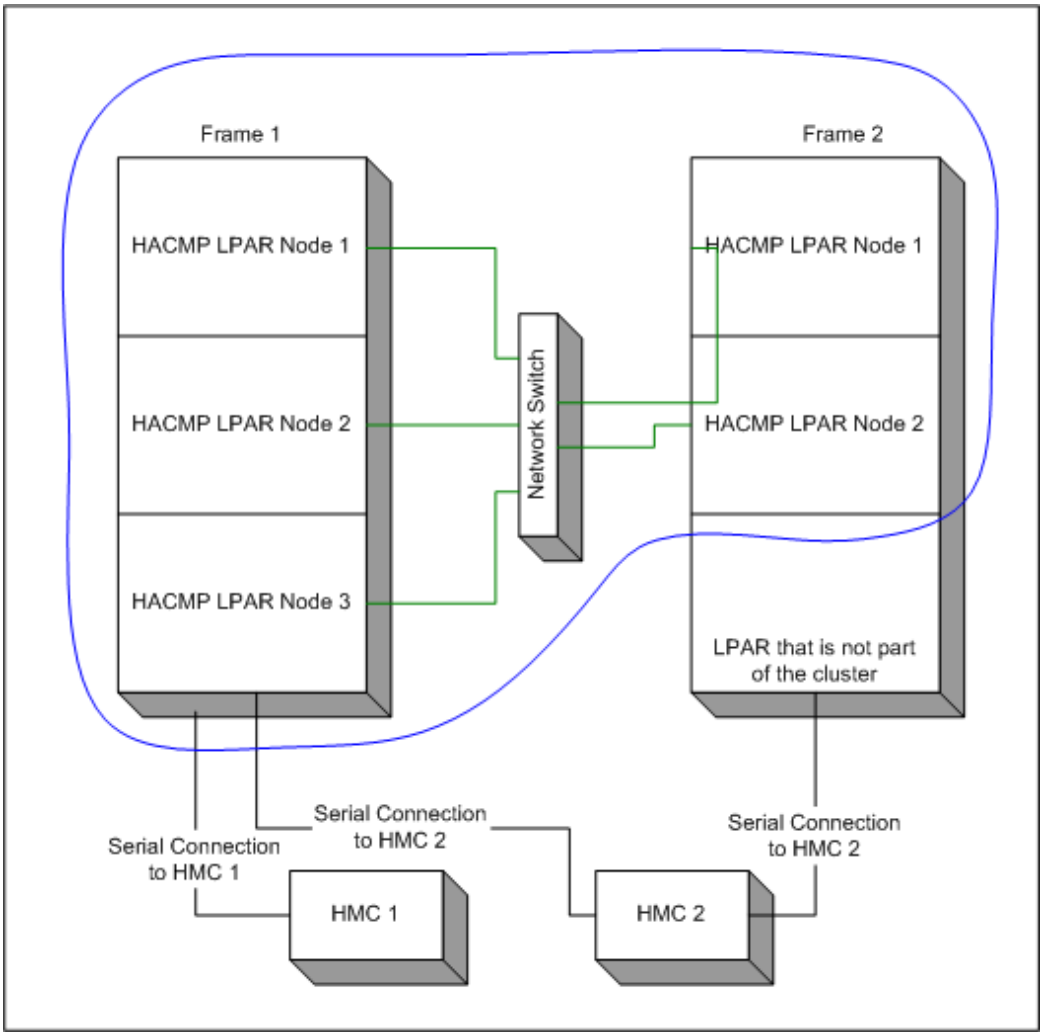
- CUoD provisioned resources. If there are *not* enough available resources in the free pool that can be allocated through DLPAR, the CUoD function provides additional resources to the standby node, should the application require more memory or CPU.

For information on how to plan CUoD in HACMP, see Planning for CUoD and DLPAR.

For information on how to configure CUoD in HACMP, see Configuring CUoD in HACMP.

A Typical HACMP Cluster to Use with the CUoD Function

You can configure an HACMP cluster within one or more pSeries servers, using two or more logical partitions. You can also configure a cluster on a subset of LPARs within one frame. Or, the cluster can use partitions from two or more frames, where the nodes can be defined as a subset of LPARs from one frame and a subset of LPARs from another frame, all connected to one or more HMCs. The following figure illustrates a typical two-frame configuration:



One of the HACMP Cluster Configurations with LPARs

Terminology for Resource Types and Memory Allocation

The following terms are used in this guide. They help you to distinguish between different types of resources allocation that can occur in an HACMP cluster that uses DLPAR and CUoD functions.

- **Total amount of resources, or permanent resources**. The number of CPUs and the amount of memory that are physically available for use by all LPARs on the frame. This amount includes all *permanent*, or *paid for* resources and may also include those CUoD resources that have already been paid for.
- **Free pool**. The number of CPUs and the amount of memory that can be dynamically allocated by HACMP through HMC to the LPAR, should the LPAR require additional resources. The free pool is the difference of the total amount of resources on the frame minus the resources that are currently used by the LPARs.

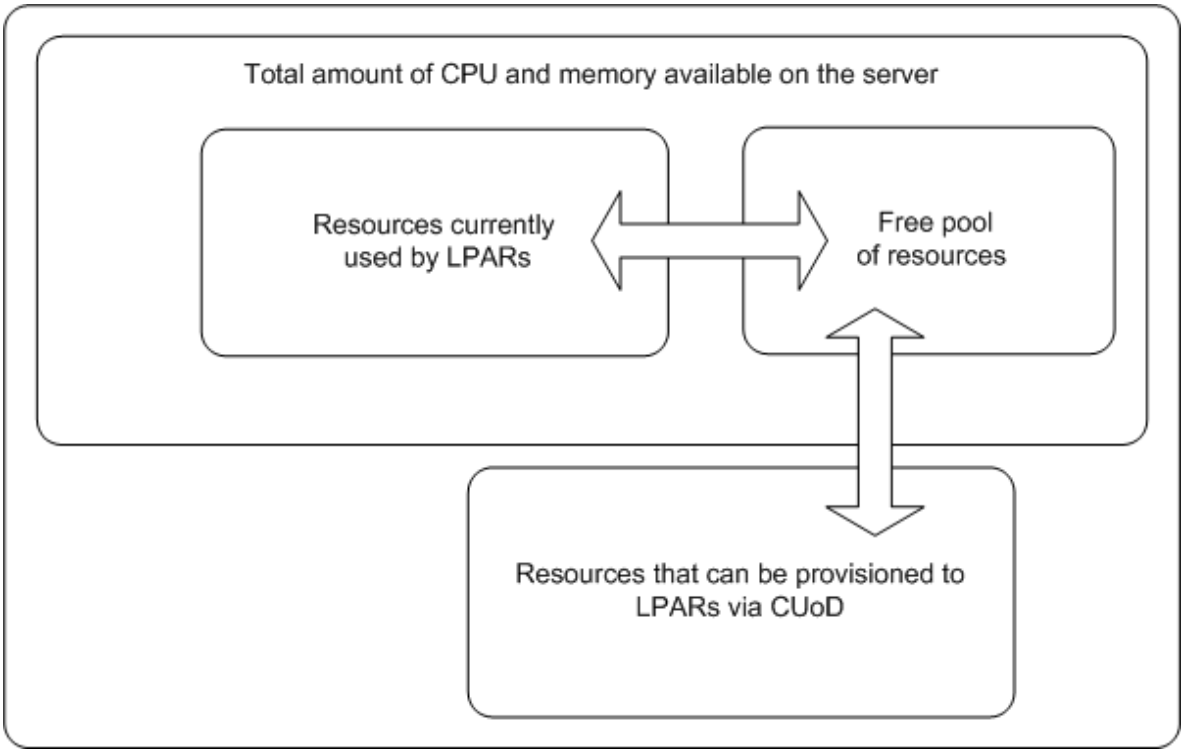
  Note: The free pool includes resources on a particular frame only. For instance, if a cluster is configured with LPARs that reside on frames A and B, HACMP does *not* request resources from a pool on frame B for an LPAR that resides on frame A.

- **CUoD pool**. The number of CPUs and the amount of memory that can be allocated by HACMP using the CUoD license, should the LPAR require additional resources. The CUoD pool depends on the type of CUoD license you have.

  Note: Note: The CUoD pool includes resources on a particular frame only.

- **LPAR minimum amount**. The minimum amount (or quantity) of a resource, such as CPU or memory, that an LPAR requires to be brought online or started. The LPAR does *not* start unless it meets the specified LPAR minimum. When DLPAR operations are performed between LPARs, the amount of resources removed from an LPAR cannot go below this value. This value is set on the HMC and is *not* modified by HACMP. Use the lshwres command on the HMC to verify this value.
- **LPAR desired amount**. The desired amount of a resource that an LPAR acquires when it starts, if the resources are available. This value is set on the HMC and is *not* modified by HACMP. Use the lshwres command on the HMC to verify this value.
- **LPAR maximum amount**. The maximum amount (or quantity) of a resource that an LPAR can acquire. When DLPAR operations are performed, the amount of resources added to an LPAR cannot go above this value. This value is set on the HMC and is *not* modified by HACMP. Use the **lshwres** command on the HMC to verify this value.

The following figure illustrates the relationship between the total amount of memory and resources on the frame (server), the free pool, and the resources that could be obtained through CUoD:



Resources currently used by LPARS, resources in the free pool and resources available through CUoD

Planning for CUoD and DLPAR

If you plan to utilize the DLPAR and CUoD functions in an HACMP cluster, it is assumed that:

- You have already planned and allocated resources to the LPARs through the HMC
- You are familiar with the types of Capacity on Demand licenses that are available.

For more information on licenses, see Configuring CUoD in HACMP in this appendix. Also refer to the Related Documentation section for a list of IBM guides that help with initial planning of LPARs and CUoD before using CUoD in HACMP.

Software and Hardware Requirements for CUoD and DLPAR

To use the CUoD and DLPAR functions in HACMP, all LPAR nodes in the cluster should have the following installed:

- AIX 5L v. 5.2 or v.5.3
- HACMP 5.3 or greater
- RSCT2.3.3.1
- HMC 3 version 2.6
- HMC build level/firmware 20040113.1 or greater
- OpenSSH 3.4p1 or greater.

Planning Requirements

Planning for Capacity on Demand in the HACMP cluster requires performing the following steps for each LPAR that serves as a cluster node:

- Obtain the LPAR resources information and resource group policies information:
  - How much memory and resources the applications that are supported by your cluster require when they run on their regular hosting nodes. Under normal running conditions, check how much memory and what number of CPUs each application uses to run with optimum performance on the LPAR node on which its resource group resides normally (home node for the resource group).
  - The startup, fallover and fallback policies of the resource group that contains the application server. Use the **clRGinfo** command. This identifies the LPAR node to which the resource group will fall over, in case of a failure.
  - How much memory and what number of CPUs are allocated to the LPAR node on which the resource group will fall over, should a failure occur. This LPAR node is referred to as a standby node. With these numbers in mind, consider whether the application's performance would be impaired on the standby node, if running with fewer resources.
  - Check the existing values for the LPAR minimums, LPAR maximums and LPAR desired amounts (resources and memory) specified. Use the **lshwres** command on the standby node.
- Estimate the resources the application will require:
  - The *minimum amount of resources* that the standby node requires to be allocated. Note that the CUoD resources will be used *in addition* to the existing LPAR resources currently available in the free pool on the frame through dynamic allocation of LPARs, and only if you explicitly tell HACMP to use them.
  - The *desired amount of resources* that the standby node would need to obtain through DLPAR or CUoD so that the application can run with the performance equivalent to its performance on the home node.

    In other words, for each standby node that can be hosting a resource group, estimate the *desired amount of resources* (memory and CPU) that this node requires to be allocated so that the application runs successfully. Note that the CUoD resources will be used *in addition* to the existing LPAR resources currently available in the free pool on the frame through dynamic allocation of LPARs, and only if you explicitly tell HACMP to use them.

    These minimum and desired amounts for memory and CPUs are the ones you will specify to HACMP. HACMP verifies that these amounts are contained within the boundaries of LPAR maximums that are set outside of HACMP for each LPAR node.
- Revise existing pre-and post-event scripts that were used to allocated DLPAR resources.

  If you were using LPAR nodes in your cluster before utilizing the CUoD and HACMP integration function, you may need to revise and rewrite your existing pre- and post-event scripts. For more information on this subject, see Using Pre- and Post-Event Scripts.

Types of CUoD Licenses

The following table describes the types of Capacity Upgrade on Demand (CUoD) licenses that are available. It also indicates whether HACMP allows the use of a particular license:

License Type	Description	Supported by HACMP	Comments
On/Off	**CPU**: Allows you to start and stop using processors as needs change. **Memory**: *not* allowed.	**CPU**: Yes **Memory**: N/A	HACMP does *not* manage licenses. The resources remain allocated to an LPAR until HACMP releases them through a DLPAR operation, or until you release them dynamically outside of HACMP. If the LPAR node goes down outside of HACMP, the CUoD resources are also released. For more information see Stopping LPAR Nodes.
Trial	**CPU and Memory**: The resources are activated for a single period of 30 consecutive days. If your system was ordered with CUoD features and they have *not* yet been activated, you can turn the features on for a one-time trial period. With the trial capability, you can gauge how much capacity you might need in the future, if you decide to permanently activate the resources you need.	**CPU**: Yes **Memory:** Yes	HACMP activates and deactivates trial CUoD resources. **Note:** Once the resources are deactivated, the trial license is used and cannot be reactivated.

Allocation and Release of Resources when Using Different Licenses

The trial and on/off types of licenses differ in how they allocate and release resources during the initial application startup (when resources are requested), and during the fallover of the resource group containing the application to another LPAR node.

To summarize, the differences are as follows:

When HACMP determines that a particular type of trial CUoD resource is needed, it activates the entire amount of that resource. This is done because only one activation instance of a particular trial CUoD resource is allowed. Note that even though HACMP activates all trial CUoD resources, only what is needed for the application is allocated to the LPAR node. The remaining resources are left in the free pool. For instance, if you have 16 GB of trial CUoD memory available, and request 2 GB, all 16 GB will be put into the free pool, but the application will acquire 2 GB.

When, during an application fallover or when stopping the application server HACMP places the allocated trial CUoD resources back into the free pool, these resources are de-activated (that is, moved to the CUoD pool) only if the trial time has expired, otherwise they remain in the free pool.

In detail, the resources are allocated and released as follows:

Startup and Fallover when Using the Trial License

Application startup. If an application server needs to allocate any additional memory or CPU resources from the CUoD pool, then *all* available memory or CPU resources from the CUoD pool (that is, *all* resources from the trial license) are allocated to the free pool and then the necessary resources from that amount are allocated to the LPAR node.

Application Fallover. If an application server has allocated any CPU or memory resources from the CUoD pool, upon fallover all CPU or memory resources are released into the free pool. Upon fallover to another LPAR node, the application then acquires only the required memory and CPU resources for the application server to come online. No CPU or memory resources are released back into the CUoD pool until the trial time expires.

Startup and Fallover when Using the On/Off License

Application Startup. If an application server needs to allocate CPU resources from the CUoD pool, *only the needed amount* of resources is allocated to the free pool and then to the LPAR node, in order for an application server to come online.

Application Fallover. If an application server has allocated CPU resources, upon fallover resources that were initially allocated from the CUoD pool will be released into the CUoD pool. Similarly, resources that were initially allocated from the free pool will be released into the free pool. Upon fallover, the application on the fallover node initiates the allocation again and allocates the required resources from the free and CUoD pools in order for the application server to come online.

Configuring CUoD in HACMP

This section contains the following:

Overview

You can configure resources that use DLPAR and CUoD operations to make these resources available to applications, if needed, for instance, when applications fall over to standby LPAR nodes.

At a high level, to enable the use of dynamically allocated LPARs and CUoD in HACMP:

1. Configure application servers.

For each application server, check its resource group policies and identify to which LPAR nodes the group (and its application) could potentially fall over, should fallover occur in the cluster.

2. For each LPAR node, establish a communication path between the node and one or more Hardware Management Consoles (HMCs). This includes configuring IP addresses HACMP will use to communicate with the HMC(s), and the Managed System name.

   If no connection to the HMC is configured for this node, HACMP assumes that this node is *not* DLPAR-capable. This allows you to have a cluster configuration with non-LPAR backup nodes (along with LPAR nodes).

3. Confirm that you want to use CUoD if the amount of DLPAR resources in the free pool is *not* sufficient. (Note that this requires accepting the license first and may result in additional charges). By default, this option in SMIT is set to **No**.

4. Configure the minimum and the desired amounts of CPU and memory that you would like to provision for your application server with the use of DLPAR function and CUoD. This task is often referred to as *application provisioning*.

Prerequisites

Prior to using the Capacity Upgrade on Demand (CUoD) function in HACMP do the following:

- Check software and hardware levels.

  Verify that the system is configured to use the required software and hardware for the DLPAR/CUoD integration. For information, see Software and Hardware Requirements for CUoD and DLPAR.

- Check the LPAR node name.

  The node name on the HMC must be the same as the AIX 5L hostname. HACMP uses the hostname to pass DLPAR commands to the HMC.

- Check what DLPAR resources you have available, and to what CUoD licenses you have access.

  HACMP does *not* tell in advance whether required resources are going to be available in all circumstances. HACMP has no control over whether you actually made the resources physically available on the pSeries frame and whether they currently remain unallocated and free. In addition, HACMP provides dynamic allocations only for CPU and Memory resources. HACMP does *not* allow dynamic changes of the I/O slots.

- Enter the license key (activation code) for CUoD.

  Obtain and enter the license key (also called the activation code) on the Hardware Management Console (HMC). Note that this may result in extra charges due to the usage of the CUoD license.

  For information on the activation code, see the IBM's *Planning Guide for Capacity Upgrade on Demand.*

- Establish secure connections to the HMC.

  Since HACMP has to securely communicate with LPAR nodes through HMCs, you must correctly install and configure SSH to allow HACMP's access to the HMC without forcing a username and password to be entered each time.

  - In the HMC's System Configuration window, put a checkmark for the option **Enable remote command execution using the SSH facility**.

- Install SSH on top of AIX 5L and generate the public and private keys necessary.

  HACMP 5.2 and up always uses the “root” user on the cluster nodes to issue SSH commands to the HMC. On the HMC system, the commands run as “hscroot” user.

  For further information on configuring SSH and remote execution, refer to the *Hardware Management Console for pSeries Installations and Operations Guide.*

Steps for Configuring DLPAR and CUoD in HACMP

To configure HACMP for DLPAR and CUoD resources:

1. Enter `smit hacmp`
2. In SMIT, select **Extended Configuration > Extended Resource Configuration > HACMP Extended Resources Configuration > Configure HACMP Applications > Configure HACMP for Dynamic LPAR and CUoD Resources** and press Enter.

   The **Configure HACMP for Dynamic LPAR and CUoD Resources** screen appears.
3. Select one of the two options:

Configure Communication Path to HMC	Select this option to establish a communication path between a node, a Managed System and one or more Hardware Management Consoles (HMCs). If a communication path is *not* established for a node, the node is considered *not* DLPAR capable.
Configure Dynamic LPAR and CUoD Resources for Applications	Select this option to configure CPU and memory resource requirements for an application server in a cluster that uses LPAR nodes.

4. Press Enter.

   Depending on which option you selected, HACMP prompts you to either configure a communication path first, or to configure resource requirements.

Configuring a Communication Path to HMC

To establish a communication path between an LPAR node and an HMC, for each LPAR node:

1. In SMIT, select **Extended Configuration > Extended Resource Configuration > HACMP Extended Resources Configuration > Configure HACMP Applications > Configure HACMP for Dynamic LPAR and CUoD Resources > Configure Communication Path to HMC > Add HMC IP Address for a Node** and press Enter.

   The **Add HMC IP Address** screen appears.
2. Enter field values as follows:

Node Name	Select a node name to associate with one or more Hardware Management Console (HMC) IP addresses and a Managed System.

HMC IP Address(es)	Enter one or more space-separated IP addresses for the HMC. If addresses are added for more than one HMC, HACMP tries to communicate with each HMC until a working communication path is found. Once the communication path is established, HACMP uses this path to execute the dynamic logical partition commands on that HMC.
Managed System Name	Enter the name of the Managed System that runs the LPAR that represents the node. The maximum length is 32 characters (do *not* enter underscores).

3. Press Enter.

   HACMP verifies that the HMC is reachable and establishes a communication path to the HMC for the node.

   Note that in some LPAR configurations outside of HACMP, LPARs could use two or more HMCs. HACMP uses only those HMCs to which the communication paths are configured. You may want to configure HACMP paths for all HMCs in the system, or only for a subset of HMCs. If HACMP successfully establishes a connection to an HMC and sends a command to it but the command fails, HACMP does *not* attempt to run this command on other HMCs for which IP labels are configured in HACMP.

Changing, Showing or Deleting Communication Path to HMC

Use the parent panel **Configure Communication Path to HMC** for changing the communication paths, and also for deleting the communication paths. Before proceeding, HACMP prompts you to select an LPAR node for which you want to change, show or delete the communication path.

Configuring Dynamic LPAR and CUoD Resources for Applications

To configure dynamic LPAR and CUoD resources, for each application server that could use DLPAR-allocated or CUoD resources:

1. In SMIT, select **Extended Configuration > Extended Resource Configuration > HACMP Extended Resources Configuration > Configure HACMP Applications > Configure HACMP for Dynamic LPAR and CUoD Resources > Configure Dynamic LPAR and CUoD Resources for Applications > Add Dynamic LPAR and CUoD Resources for Applications** and press Enter.

   A picklist of configured application servers appears.

2. Select an application server from the list and press Enter.

   The screen for specifying the resource requirements for an application server appears.

3. Enter field values as follows:

Application Server Name	The name of the application server appears here. This is the application server for which you will configure Dynamic LPAR and CUoD resource provisioning. You cannot edit values in this field.

Minimum Number of CPUs

Enter the minimum number of CPUs to acquire when the application server starts. Enter only whole numbers; you cannot enter decimal or fractional numbers in this field. The default value is 0.

To perform the application provisioning, HACMP checks how many CPUs the LPAR node currently has above its LPAR minimum value, compares this number with the minimum requested in this field and based on this, requests more CPUs, if needed.

Note that the LPAR minimum for CPUs is a starting point for HACMP's calculations. For instance, if the LPAR minimum is 2 and the LPAR currently owns only 2 CPUs, but the application server requests 3, this means that HACMP will request 3 more CPUs. HACMP attempts to allocate 3 CPUs first from the free pool on the frame and then through the CUoD function, if you allow to use CUoD.

If HACMP cannot satisfy the amount of CPUs specified in this field for the application server, HACMP takes resource group recovery actions to move the resource group with its application to another node.

Desired Number of CPUs

Enter the maximum amount of CPUs HACMP will attempt to allocate to the node before starting this application on this node. Enter only whole numbers; you cannot enter decimal or fractional numbers in this field. The default value is 0.

At configuration time, HACMP checks that this value is greater than or equal to the **Minimum Number of CPUs** (defined in the field above).

HACMP may allocate fewer CPUs if there are *not* enough available in the free pool.

Minimum Amount of Memory (in 256 MB increments)

Enter the amount of memory to acquire when the application server starts.

To perform the application provisioning, HACMP checks how much memory the LPAR node currently has above its LPAR minimum, compares this amount with the minimum requested in this field and based on this and requests more memory, if needed.

Note that the LPAR minimum for memory is a starting point for HACMP calculations. For instance, if the LPAR minimum is 256 MB and the LPAR currently owns only 256 MB and does *not* host any other application servers, but the application server requests 512 MB, this means that HACMP will request 512MB more memory. HACMP attempts to allocate the additional memory first from the free pool on the frame and then through the CUoD function, if you allow to use CUoD.

If this amount of memory is *not* satisfied, HACMP takes resource group recovery actions to move the resource group with its application to another node.

Desired Amount of Memory (in 256 MB increments)

Enter the maximum amount of memory HACMP will attempt to allocate to the node before starting this application. The default value is 0.

At configuration time, HACMP checks that this value is greater than or equal to the **Minimum Amount of Memory** (defined in the field above).

HACMP can allocate less memory if there is *not* enough available.

Use CUoD if resources are insufficient?

The default is **No**. Select **Yes** to have HACMP use Capacity Upgrade on Demand (CUoD) to obtain enough resources to fulfill the minimum amount requested. Using CoD requires a license key (activation code) to be entered on the Hardware Management Console (HMC) and may result in extra costs due to usage of the CoD license.

If you plan to use resources that will be allocated through CUoD (in addition to dynamic allocation of resources available in the free pool on the frame), the answer to this question must be **Yes**.

HACMP calculates the number of resources to be acquired through CUoD in the following way:

If the amount of resources currently on the DLPAR, plus the amount that can be allocated from the free pool does *not* meet the requested minimum for the application server, HACMP requests the additional resources through CUoD.

I agree to use CUoD resources (this might result in extra costs by using the CUoD license)	The default is **No**. Select **Yes** to acknowledge that you understand that there might be extra costs involved when using CUoD. HACMP logs the answer to the **syslog** and **smit.log** files.

4. Press Enter.

   When the application requires additional resources to be allocated on this node, HACMP performs its calculations to see whether it needs to request only the DLPAR resources from the free pool on the frame and whether that would already satisfy the requirement, or if CUoD resources are also needed for the application server. After that, HACMP proceeds with requesting the desired amounts of memory and numbers of CPU, if you selected to use them.

   During verification, HACMP ensures that the entered values are below LPAR maximum values for memory and CPU. Otherwise HACMP issues an error, stating these requirements.

   HACMP also verifies that the total of required resources for *ALL* application servers that can run concurrently on the LPAR is less than the LPAR maximum. If this requirement is *not* met, HACMP issues a warning. Note that this scenario can happen upon subsequent fallovers. That is, if the LPAR node is already hosting application servers that require DLPAR and CUoD resources, then upon acquiring yet another application server, it is possible that the LPAR cannot acquire any additional resources beyond its LPAR maximum. HACMP verifies this case and issues a warning.

Changing Dynamic LPAR and CUoD Resources for Applications

To change or show dynamic LPAR and CUoD resources:

1. In SMIT, select **Extended Configuration > Extended Resource Configuration > HACMP Extended Resources Configuration > Configure HACMP Applications > Configure HACMP for Dynamic LPAR and CUoD Resources > Configure Dynamic LPAR and CUoD Resources for Applications > Change/Show Dynamic LPAR and CUoD Resources for Applications** and press Enter.

   A picklist of configured application servers appears.

2. Select an application server from the list and press Enter.

   The screen displays the previously entered values for the application server's minimum and desired amounts. Note that each time you request CUoD resources, you must select Yes in the appropriate fields to allow HACMP to proceed with using CUoD.

3. Change field values as follows:

Application Server Name	The name of the application server appears here. This is the application server for which you will configure Dynamic LPAR and CUoD resource provisioning. You cannot edit values in this field.

Minimum Number of CPUs

Enter the minimum number of CPUs to acquire when the application server starts. Enter only whole numbers; you cannot enter decimal or fractional numbers in this field. The default value is 0.

To perform the application provisioning, HACMP checks how many CPUs the LPAR node currently has above its LPAR minimum, compares this number with the minimum requested in this field and based on this, requests more CPUs, if needed.

Note that the LPAR minimum for CPUs is a starting point for HACMP's calculations. For instance, if the LPAR minimum is 2 and the LPAR currently owns only 2 CPUs, but the application server requests 3, this means that HACMP will request 3 more CPUs. HACMP attempts to allocate 3 CPUs first from the free pool on the frame and then through the CUoD function, if you allow to use CUoD.

If HACMP cannot satisfy the amount of CPUs specified in this field for the application server, HACMP takes resource group recovery actions to move the resource group with its application to another node.

Desired Number of CPUs

Enter the maximum amount of CPUs HACMP will attempt to allocate to the node before starting this application on this node. Enter only whole numbers; you cannot enter decimal or fractional numbers in this field. The default value is 0.

At configuration time, HACMP checks that this value is greater than or equal to the **Minimum Number of CPUs** (defined in the field above).

HACMP may allocate fewer CPUs if there are *not* enough available in the free pool.

Minimum Amount of Memory (in 256 MB increments)

Enter the amount of memory to acquire when the application server starts.

To perform the application provisioning, HACMP checks how much memory the LPAR node currently has above its LPAR minimum, compares this amount with the minimum requested in this field and based on this, and requests more memory, if needed.

Note that the LPAR minimum for memory is a starting point for HACMP calculations. For instance, if the LPAR minimum is 256 MB and the LPAR currently owns only 256 MB and does *not* host any other application servers, but the application server requests 512 MB, this means that HACMP will request 512MB more memory. HACMP attempts to allocate the additional memory first from the free pool on the frame and then through the CUoD function, if you allow to use CUoD.

If this amount of memory is *not* satisfied, HACMP takes resource group recovery actions to move the resource group with its application to another node.

Desired Amount of Memory (in 256 MB increments)

Enter the maximum amount of memory HACMP will attempt to allocate to the node before starting this application. The default value is 0.

At configuration time, HACMP checks that this value is greater than or equal to the **Minimum Amount of Memory** (defined in the field above).

HACMP can allocate less memory if there is *not* enough available.

Use CUoD if resources are insufficient?

The default is **No**. Select **Yes** to have HACMP use Capacity Upgrade on Demand (CUoD) to obtain enough resources to fulfill the minimum amount requested. Using CoD requires a license key (activation code) to be entered on the Hardware Management Console (HMC) and may result in extra costs due to usage of the CoD license.

If you plan to use resources that will be allocated through CUoD (in addition to dynamic allocation of resources available in the free pool on the frame), the answer to this question must be **Yes**.

HACMP calculates the number of resources to be acquired through CUoD in the following way:

If the amount of resources currently on the DLPAR, plus the amount that can be allocated from the free pool does *not* meet the requested minimum for the application server, HACMP requests the additional resources through CUoD.

I agree to use CUoD resources (this might result in extra costs by using the CUoD license)	The default is **No**. Select **Yes** to acknowledge that you understand that there might be extra costs involved when using CUoD. HACMP logs the answer to the **syslog** and **smit.log** files.

4. Press Enter.

When the application requires additional resources to be allocated on this node, HACMP performs its calculations to see whether it needs to request only the DLPAR resources from the free pool on the frame and whether that would already satisfy the requirement, or if CUoD resources are also needed for the application server. After that, HACMP proceeds with requesting the desired amounts of memory and numbers of CPU, if you selected to use them.

Deleting Dynamic LPAR and CUoD Resources for Applications

Use the parent panel **Configure Dynamic LPAR and CUoD Resources for Applications** for for removing the resource requirements for application servers. The **Remove Dynamic LPAR and CUoD Resources for Applications** screen prompts you to select the application server, and lets you remove the application resource provisioning information.

If you delete the application server, HACMP also deletes the application provisioning information for it.

Changing the DLPAR and CUoD Resources Dynamically

You can change the DLPAR and CUoD resource requirements for application servers without stopping the cluster services. Synchronize the cluster after making the changes.

The new configuration is *not* reflected until the next event that causes the application (hence the resource group) to be released and reacquired on another node. In other words, a change in the resource requirements for CPUs, memory or both does *not* cause the recalculation of the DLPAR resources. HACMP does *not* stop and restart application servers solely for the purpose of making the application provisioning changes.

If *another* dynamic reconfiguration change causes the resource groups to be released and reacquired, the new resource requirements for DLPAR and CUoD are used at the end of this dynamic reconfiguration event.

How Application Provisioning Works in HACMP

This section describes the flow of actions in the HACMP cluster, if the application provisioning function through DLPAR and CUoD is configured. It also includes several examples that illustrate how resources are allocated, depending on different resource requirements.

In addition, the section provides some recommendations on using pre-and post-scripts.

Overview

When you configure an LPAR on the HMC (outside of HACMP), you provide LPAR minimum and LPAR maximum values for the number of CPUs and amount of memory. You can obtain these values by running the commands on the HMC. The stated minimum values of the resources must be available when an LPAR node starts. If more resources are available in the

free pool on the frame, an LPAR can allocate up to the stated desired values. During dynamic allocation operations, the system does *not* allow that the values for CPU and memory go below the minimum or above the maximum amounts specified for the LPAR.

HACMP obtains the LPAR minimums and LPAR maximums amounts and uses them to allocate and release CPU and memory when application servers are started and stopped on the LPAR node.

HACMP requests the DLPAR resource allocation on the HMC before the application servers are started, and releases the resources after the application servers are stopped. The Cluster Manager waits for the completion of these events before continuing the event processing in the cluster.

HACMP handles the resource allocation and release for application servers serially, regardless if the resource groups are processed in parallel. This minimizes conflicts between application servers trying to allocate or release the same CPU or memory resources. Therefore, you must carefully configure the cluster to properly handle all CPU and memory requests on an LPAR.

These considerations are important:

- Once HACMP has acquired additional resources for the application server, when the application server moves again to another node, HACMP releases only those resources that are no longer necessary to support this application on the node.
- HACMP does *not* start and stop LPAR nodes.

Acquiring DLPAR and CUoD Resources

If you configure an application server that requires a minimum and a desired amount of resources (CPU or memory), HACMP determines if additional resources need to be allocated for the node and allocates them if possible.

In general, HACMP tries to allocate as many resources as possible to meet the desired amount for the application, and uses CUoD, if allowed, to do this.

The LPAR Node has the LPAR Minimum

If the node owns *only* the minimum amount of resources, HACMP requests additional resources through DLPAR and CUoD.

In general, HACMP starts counting the extra resources required for the application from the minimum amount. That is, the minimum resources are retained for the node's overhead operations, and are *not* utilized to host an application.

The LPAR Node has Enough Resources to Host an Application

The LPAR node that is about to host an application may already contain enough resources (in addition to the LPAR minimum) to meet the desired amount of resources for this application.

In this case, HACMP does *not* allocate any additional resources and the application can be successfully started on the LPAR node. HACMP also calculates that the node has enough resources for this application in addition to hosting all other application servers that may be currently running on the node.

Resources Requested from the Free Pool and from the CUoD Pool

If the amount of resources in the free pool is insufficient to satisfy the total amount requested for allocation (minimum requirements for one or more applications), HACMP requests resources from CUoD.

If HACMP meets the requirement for a minimum amount of resources for the application server, application server processing continues. Application server processing continues even if the total desired resources (for one or more applications) have *not* been met or are only partially met. In general, HACMP attempts to acquire *up to the desired amount* of resources requested for an application.

If the amount of resources is insufficient to host an application, HACMP starts resource group recovery actions to move the resource group to another node.

The Minimum Amount Requested for an Application Cannot be Satisfied

In some cases, even after HACMP requests to use resources from the CUoD pool, the amount of resources it can allocate is less than the minimum amount specified for an application.

If the amount of resources is still insufficient to host an application, HACMP starts resource group recovery actions to move the resource group to another node.

The LPAR node is Hosting Application Servers

In all cases, HACMP checks whether the node is already hosting application servers that required application provisioning, and that the LPAR maximum for the node is *not* exceeded:

- Upon subsequent fallovers, HACMP checks if the minimum amount of requested resources for yet another application server plus the amount of resources already allocated to applications residing on the node exceeds the LPAR maximum.
- In this case, HACMP attempts resource group recovery actions to move the resource group to another LPAR. Note that when you configure the DLPAR and CUoD requirements for this application server, then during cluster verification, HACMP warns you if the total number of resources requested for all applications exceeds the LPAR maximum.

Allocation of Resources in a Cluster With Multiple Applications

If you have multiple applications in different resource groups in the cluster with LPAR nodes, and more than one application is configured to potentially request additional resources through the DLPAR and CUoD function, the resource allocation in the cluster becomes more complex.

Based on the resource group processing order, some resource groups (hence the applications) might *not* be started. See Example 2: Failure to Allocate CPUs due to Resource Group Processing Order.

In general, to better understand how HACMP allocates resources in different scenarios, see Examples of using DLPAR and CUoD Resources.

Releasing DLPAR and CUoD Resources

When the application server is stopped on the LPAR node (the resource group moves to another node), HACMP releases only those resources that are no longer necessary to support this application server on the node. The resources are released to the free pool on the frame.

HACMP first releases the DLPAR or CUoD resources it acquired last. This implies that the CUoD resources may *not* always be released before the dynamic LPAR resources are released.

The free pool is limited to the single frame only. That is, for clusters configured on two frames, HACMP does *not* request resources from the second frame for an LPAR node residing on the first frame.

Also, if LPAR 1 releases an application that puts some DLPAR resources into free pool, LPAR 2, which is using the CUoD resources, does *not* make any attempt to release its CUoD resources and acquire the free DLPAR resources.

Stopping LPAR Nodes

When the Cluster Manager is forced down on an LPAR node, and that LPAR is subsequently shutdown (outside of HACMP), the CPU and memory resources are released (*not* by HACMP) and become available for other resource groups running on other LPARs. HACMP does *not* track CPU and memory resources that were allocated to the LPAR and does *not* retain them for use when the LPAR node rejoins the cluster.

Note: If you are using the On/Off license for CUoD resources, and the LPAR node is shutdown (outside of HACMP), the CUoD resources are released (*not* by HACMP) to the free pool, but the On/Off license continues to be turned on. You may need to manually turn off the licence for the CUoD resources that are now in the free pool. (This ensures that you do *not* pay for resources that are *not* being currently used).

If the LPAR is *not* stopped after the Cluster Manager is forced down on the node, the CPU and memory resources remain allocated to the LPAR for use when the LPAR rejoins the cluster.

Examples of using DLPAR and CUoD Resources

The following examples show CPU allocation and release. (Memory allocation process is similar).

It is important to remember that once HACMP acquires additional resources for an application server, when the server moves again to another node, it takes the resources with it, that is, the LPAR node releases all the additional resources it acquired, and remains with just the minimum.

The configuration is an 8 CPU frame, with a two-node (each an LPAR) cluster. There are 2 CPUs available in the CUoD pool, that is through the CUoD activations. The nodes have the following characteristics:

Node Name	LPAR Minimum	LPAR Maximum
Node1	1	9
Node2	1	5

The following application servers are defined in separate resource groups:

Application server name	CPU Desired	CPU Minimum	Allow to Use CUoD?
AS1	1	1	Yes
AS2	2	2	No
AS3	4	4	No

Example 1: No CPUs Are Allocated at Application Server Start, some CPUs are Released at Server Stop

Current configuration settings:

- Node1 has 3 CPUs allocated.
- Node2 has 1 CPU allocated.
- The free pool has 4 CPUs.

HACMP starts application servers as follows:

- Node1 starts AS2, no CPUs are allocated to meet the requirement of 3 CPUs. (3 CPUs is equal to the sum on Node1's LPAR minimum of 1 plus AS2 desired amount of 2).
- Node1 stops AS2. 2 CPUs are released, leaving 1 CPU, the minimum requirement. (Since no other application servers are running, the only requirement is Node1 LPAR minimum of 1).

Example 2: Failure to Allocate CPUs due to Resource Group Processing Order

Current configuration settings:

- Node1 has 3 CPUs allocated.
- Node2 has 1 CPU allocated.
- The free pool has 4 CPUs.

HACMP starts application servers as follows:

- Node1 starts AS1, no CPUs are allocated since the requirement of 2 is met.

  Node1 starts AS3, 3 CPUs are allocated to meet the requirement of 6. There is now 1 CPU in the free pool.
- Node1 attempts to start AS2. After Node 1 has acquired AS1 and AS3, the total amount of CPUs Node1 must now own to satisfy these requirements is 6, which is the sum of Node1 LPAR minimum of 1 plus AS1 desired amount of 1 plus AS3 desired amount of 4.

  Since AS2 minimum amount is 2, in order to acquire AS2, Node1 needs to allocate 2 more CPUs, but there is only 1 CPU left in the free pool and it does *not* meet the minimum requirement of 2 CPUs for AS2. The resource group with AS2 goes into error state since there is only 1 CPU in the free pool and CUoD use is *not* allowed.

Example 3: Successful CUoD Resources Allocation and Release

Current configuration settings:

- Node1 has 3 CPUs allocated.
- Node2 has 1 CPU allocated.
- The free pool has 4 CPUs.

HACMP starts application servers as follows:

- Node1 starts AS3, 2 CPUs are allocated to meet the requirement of 5.
- Node1 starts AS2, 2 CPUs are allocated to meet the requirement of 7. There are now no CPUs in the free pool.
- Node1 starts AS1, 1 CPU is taken from CUoD and allocated to meet the requirement of 8.
- Node1 stops AS3, 4 CPUs are released and 1 of those CPUs is put back into the CUoD pool.

Example 4: Resource Group Failure (the Minimum for the Server is not Met, but the LPAR Maximum for the Node is Reached)

Current configuration settings:

- Node1 has 1 CPU allocated.
- Node2 has 1 CPU allocated.

- The free pool has 6 CPUs.

HACMP starts application servers as follows:

- Node2 starts AS3, 4 CPUs are allocated to meet the requirement of 5. There are now 2 CPUs in the free pool.
- Node2 attempts to start AS2, but AS2 goes into error state since the LPAR maximum for Node2 is 5 and Node2 cannot acquire more CPUs.

Example 5: Resource Group Fallover

Current configuration settings:

- Node1 has 3 CPUs allocated.
- Node2 has 1 CPU allocated.
- The Free pool has 4 CPUs.

HACMP starts application servers as follows:

- Node1 starts AS2, no CPUs are allocated to meet the requirement of 3.
- The resource group with AS2 falls over from Node1 to Node2.
- Node1 stops AS2. 2 CPUs are released, leaving 1 CPU on the LPAR, the minimum requirement for the node.
- Node2 start AS2, 2 CPUs are allocated to meet the requirement of 3.

Using Pre- and Post-Event Scripts

The existing pre- and post-event scripts that you were using in a cluster with LPARs (before using the CUoD integration with HACMP) may need to be modified or rewritten, if you plan to configure CUoD and DLPAR requirements in HACMP.

Keep in mind the following:

- HACMP performs all the DLPAR operations before the application servers are started, and after they are stopped. You may need to rewrite the scripts to account for this.
- Since HACMP takes care of the resource calculations, requests additional resources from the DLPAR operations and, if allowed, from CUoD, you may get rid of the portions of your scripts that do that.
- HACMP only takes into consideration the free pool on a single frame. If your cluster is configured within one frame, then modifying the scripts as stated above is sufficient.

  However, if a cluster is configured with LPAR nodes residing on two frames, you may still require the portions of the existing pre- and post-event scripts that deal with dynamically allocating resources from the free pool on one frame to the node on another frame, should the application require these resources.

Troubleshooting DLPAR and CUoD Operations in HACMP

To troubleshoot the DLPAR operations in your cluster, use the event summaries in the **hacmp.out** file and **syslog**.

HACMP logs the following information:

- All resource allocation attempts along with the outcome in the event summaries. If CUoD resources are used, a separate log entry indicates this in the event summary in **hacmp.out** and in **syslog**.
- The information about the released resources.
- Each time you use CUoD, remember to select **Yes** in the appropriate SMIT fields for application provisioning configuration. HACMP logs your replies in **syslog**.

HACMP processing may wait until the DLPAR operations are completed on a particular node. This also affects the time it takes HACMP to start a resource group. Use the event summaries to track processing.

Use the following commands on the LPAR node or on the HMC:

lshwres	View the LPAR minimum, LPAR maximum and the total amount of memory and the number of CPUs that are currently allocated to the LPAR.
lssyscfg	Verify that the LPAR node is DLPAR capable.
chhwres	Run the DLPAR operations on the HMC outside of HACMP and to manually change the LPAR minimum, LPAR minimum and LPAR desired values for the LPAR. This may be necessary if HACMP issues an error or a warning, during the verification process, if you requested to use DLPAR and CUoD resources in HACMP.
hmc_exec in clhmcexec	Logs HMC commands upon verification errors. This may be helpful if the HACMP cluster verification fails.

Notices for HACMP Administration Guide

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing
IBM Corporation
North Castle Drive
Armonk, NY 10504-1785
U.S.A.

For license inquiries regarding double-byte (DBCS) information, contact the IBM Intellectual Property Department in your country or send inquiries, in writing, to:

IBM World Trade Asia Corporation
Licensing
2-31 Roppongi 3-chome, Minato-ku
Tokyo 106, Japan

The following paragraph does not apply to the United Kingdom or any country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION “AS IS” WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions; therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Licensees of this program who wish to have information about it for the purpose of enabling: (i) the exchange of information between independently created programs and other programs (including this one) and (ii) the mutual use of the information which has been exchanged, should contact:

IBM Corporation
Dept. LRAS / Bldg. 003
11400 Burnet Road
Austin, TX 78758-3493
U.S.A.

Such information may be available, subject to appropriate terms and conditions, including in some cases, payment of a fee.

The licensed program described in this document and all licensed material available for it are provided by IBM under terms of the IBM Customer Agreement, IBM International Program License Agreement or any equivalent agreement between us.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

Index

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

R

S

T

U

V

W

XYZ