Dell EMC Host Connectivity Guide for IBM AIX

P/N 300-000-608
Rev. 59

June 2020

DELL EMC

Dell EMC
Hopkinton, Massachusetts 01748-9103
1-508-435-1000 In North America 1-866-464-7381
www.DellEMC.com

CONTENTS

FIGURES

TABLES

PREFACE

As part of an effort to improve and enhance the performance and capabilities of its product line, Dell EMC from time to time releases revisions of its hardware and software. Therefore, some functions described in this document may not be supported by all revisions of the software or hardware currently in use. For the most up-to-date information on product features, refer to your product release notes.

If a product does not function properly or does not function as described in this document, please contact your Dell EMC representative.

This guide describes the features and setup procedures for IBM AIX host interfaces to the EMC VNX series and Dell EMC VMAX3, VMAX, Symmetrix, and CLARiiON storage systems over Fibre Channel and (Symmetrix only) SCSI.

IMPORTANT

EMC recommends using AIX default settings when connecting AIX hosts to EMC storage arrays unless otherwise mentioned in this guide or other EMC product manuals.

Audience

This guide is intended for use by storage administrators, system programmers, or operators who are involved in acquiring, managing, or operating VMAX3, VMAX, Symmetrix, VNX series, or CLARiiON, and host systems.

Readers of this guide are expected to be familiar with:

- VMAX3, VMAX, Symmetrix, VNX, and CLARiiON system operation
- IBM AIX operating environment

AIX, VMAX, Symmetrix, and CLARiiON references

Unless otherwise noted:

- Any references to AIX server or AIX host include the RS/6000, RS/6000 SP, eServer pSeries, System p5, System p6, System p, JS-Series BladeCenter, and Bull OEM server.
- Any general references to Symmetrix or Symmetrix array include the VMAX3 Family, VMAX, and DMX.
- Any general references to VNX include VNX5100/5200/5300/ 5400/5500/5600/5700/5800/7500/7600/8000.
- Any general references to CLARiiON or CLARiiON array include the FC4700, FC4700-2, CX, CX3, and CX4.

Table 1 lists the minimum Dell EMC Enginuity™ and Dell EMC HYPERMAX requirements for VMAX and Symmetrix models.

Table 1 Dell EMC models and minimum operating system requirements

EMC model	Minimum operating system
VMAX 400K [1]	HYPERMAX 5977.250.189
VMAX 200K [1]	HYPERMAX 5977.250.189
VMAX 100K [1]	HYPERMAX 5977.250.189
VMAX 40K	Enginuity 5876.82.57
VMAX 20K	Enginuity 5876.82.57
VMAX 10K (Systems with SN xxx987xxxx)	Enginuity 5876.159.102
VMAX	Enginuity 5876.82.57
VMAX 10K (Systems with SN xxx959xxxx)	Enginuity 5876.82.57
VMAXe®	Enginuity 5876.82.57
Symmetrix DMX-4	Enginuity 5773.79.58
Symmetrix DMX-3	Enginuity 5773.79.58

1. VMAX 400K, VMAX 200K, and VMAX 100K are also referred to as the VMAX3™ Family.

Related documentation

For Dell EMC documentation, refer to Dell EMC Online Support.

For the most up-to-date information, always consult the Dell EMC Simple Support Matrices (ESM), available through Dell E-Lab Navigator (ELN).

Conventions used in this guide

Dell EMC uses the following conventions for notes and cautions.

Note: A note presents information that is important, but not hazard-related.

IMPORTANT

An important notice contains information essential to operation of the software.

Typographical Conventions

EMC uses the following type style conventions in this guide.

Normal — Used in running (nonprocedural) text for:
- Names of interface elements, such as names of windows, dialog boxes, buttons, fields, and menus
- Names of resources, attributes, pools, Boolean expressions, buttons, DQL statements, keywords, clauses, environment variables, functions, and utilities
- URLs, pathnames, filenames, directory names, computer names, links, groups, service keys, file systems, and notifications

Bold — Used in running (nonprocedural) text for names of commands, daemons, options, programs, processes, services, applications, utilities, kernels, notifications, system calls, and man pages

	Used in procedures for: • Names of interface elements, such as names of windows, dialog boxes, buttons, fields, and menus • What the user specifically selects, clicks, presses, or types
Italic	Used in all text (including procedures) for: • Full titles of publications referenced in text • Emphasis, for example, a new term • Variables
`Courier`	Used for: • System output, such as an error message or script • URLs, complete paths, filenames, prompts, and syntax when shown outside of running text
`Courier bold`	Used for specific user input, such as commands
`Courier italic`	Used in procedures for: • Variables on the command line • User input variables
< >	Angle brackets enclose parameter or variable values supplied by the user
[ ]	Square brackets enclose optional values
|	Vertical bar indicates alternate selections — the bar means “or”
{ }	Braces enclose content that the user must specify, such as x or y or z
...	Ellipses indicate nonessential information omitted from the example

Where to get help

Dell EMC support, product, and licensing information can be obtained as follows.

Dell EMC support, product, and licensing information can be obtained on Dell EMC Online Support.

Note: To open a service request through the Dell EMC Online Support site, you must have a valid support agreement. Contact your EMC sales representative for details about obtaining a valid support agreement or to answer any questions about your account.

Product information

For documentation, release notes, software updates, or for information about Dell EMC products, licensing, and service, go to Dell EMC Online Support.

Technical support

Dell EMC offers a variety of support options.

Support by Product — Dell EMC offers consolidated, product-specific information on the Web at the Dell EMC Online Support By Product page.

The Support by Product web pages offer quick links to Documentation, White Papers, Advisories (such as frequently used Knowledgebase articles), and Downloads, as well as more dynamic content, such as presentations, discussion, relevant Customer Support Forum entries, and a link to Dell EMC Live Chat.

Dell EMC Live Chat — Open a Chat or instant message session with a Dell EMC Support Engineer.

eLicensing support

To activate your entitlements and obtain your Symmetrix license files, visit the Service Center on Dell EMC Online Support, as directed on your License Authorization Code (LAC) letter that was e-mailed to you.

For help with missing or incorrect entitlements after activation (that is, expected functionality remains unavailable because it is not licensed), contact your Dell EMC Account Representative or Authorized Reseller.

For help with any errors applying license files through Solutions Enabler, contact the Dell EMC Customer Support Center.

If you are missing a LAC letter, or require further instructions on activating your licenses through the Online Support site, contact EMC's worldwide Licensing team at licensing@emc.com or call:

- North America, Latin America, APJK, Australia, New Zealand: SVC4EMC (800-782-4362) and follow the voice prompts.
- EMEA: +353 (0) 21 4879862 and follow the voice prompts.

We'd like to hear from you!

Your suggestions will help us continue to improve the accuracy, organization, and overall quality of the user publications. Send your opinions of this document to:

techpubcomments@emc.com

Your feedback on our TechBooks is important to us! We want our books to be as helpful and relevant as possible. Send us your comments, opinions, and thoughts on this or any other TechBook to:

TechBooks@emc.com

PART 1

General AIX Information

Part 1 includes:

CHAPTER 1

General AIX Information

This chapter provides general information about AIX hosts and the AIX operating system.

Note: Any general references to Symmetrix or Symmetrix array include the VMAX3 Family, VMAX, and DMX.

Features of the AIX operating system

Following are brief discussions of unique system management features of the operating system.

Logical Volume Manager

The Logical Volume Manager (LVM) maintains the hierarchy of logical structures that manage disk storage. Disks are defined within this hierarchy as physical volumes. Every physical volume in use belongs to a volume group. Within each volume group, one or more logical volumes of information are defined. Data on logical volumes appears to be contiguous to the user, but can be discontiguous on the physical volume. This allows file systems, paging space, and other logical volumes to be resized or relocated, span multiple physical volumes, and have their contents replicated for greater flexibility and availability.

JFS and JFS2

JFS2 (enhanced journaled file system), introduced in AIX 5L for POWER Version 5.1, provides the capability to store much larger files than the existing journaled file system (JFS). Customers can choose to implement either JFS (the recommended file system for 32-bit environments) or JFS2 (which offers 64-bit functionality).

Table 2 JFS2 and JFS functions

Functions	JFS2	JFS
Fragments/block size	512–4096 block sizes	512–4096 fragments
Maximum file system size	16 TB	1 TB
Maximum file size	16 TB	64 GB
Number of i-nodes	Dynamic; limited by disk space	Fixed; set at file system creation
Directory organization	B-tree	Linear
Compression	No	Yes
Quotas	No	Yes

Note: The maximum file size and maximum file system size are limited to 1 TB (terabyte) when used with the 32-bit kernel.

As Table 2 on page 18 shows the JFS divides disk space into allocation units called *fragments*, while the JFS2 segments disk space into *blocks*. The objective is the same—to efficiently store data.

JFS fragments are smaller than the default disk allocation size of 4096 bytes. Fragments minimize wasted disk space by more efficiently storing the data in a file or directory's partial logical blocks.

FS2 supports multiple file system block sizes of 512, 1024, 2048, and 4096. Smaller block sizes minimize wasted disk space by more efficiently storing the data in a file or

directory's partial logical blocks. Smaller block sizes also result in additional operational overhead. The block size for a JFS2 is specified during its creation. Different file systems can have different block sizes; however, only one block size can be used within a single file system.

System Resource Controller (SRC)

The System Resource Controller (SRC) provides a set of commands and subroutines for creating and controlling subsystems. It provides a mechanism to control subsystem processes by using a command-line or C interface. This allows you to start, stop, and collect status information on subsystem processes with shell scripts, commands, or user-written programs.

Object Data Manager

The Object Data Manager (ODM) is a data manager intended for the storage of system data. Many system management functions use the ODM database. Information used in many commands and SMIT functions is stored and maintained as objects with associated characteristics. System data managed by ODM includes: Device configuration information, Display information for SMIT (menus, selectors, and dialogs), Vital product data for installation and update procedures, Communications configuration information and System resource information.

Software Vital Product Data

Certain information about software products and their installable options is maintained in the Software Vital Product Data (SWVPD) database. The SWVPD consists of a set of commands and Object Database Manager (ODM) object classes for the maintenance of software product information. The SWVPD commands are provided for the user to query (`lslpp`) and verify (`lppchk`) installed software products. The ODM object classes define the scope and format of the software product information that is maintained. The **`installp`** command uses the ODM to maintain the following information in the SWVPD database: Name of the installed software product, Version of the software product, Release level of the software product, which indicates changes to the external programming interface of the software product, Modification level of the software product, which indicates changes that do not affect the external programming interface of the software product, Fix level of the software product, which indicates small updates that are to be built into a regular modification level at a later time, Fix identification field, Names, checksums, and sizes of the files that make up the software product or option, and Installation state of the software product: applying, applied, committing, committed, rejecting, or broken.

Workload Manager (WLM)

Workload Manager (WLM) lets you create different classes of service for jobs, and specify attributes for those classes. These attributes specify minimum and maximum amounts of CPU, physical memory, and disk I/O throughput to be allocated to a class. WLM then assigns jobs automatically to classes using class assignment rules provided by a system administrator. These assignment rules are based on the values of a set of attributes for a process. Either the system administrator or a privileged user can also manually assign jobs to classes, overriding the automatic assignment.

Logical volume overview

Each independent disk called a physical volume (PV) has a name, such as `/dev/hdisk0`. Every physical volume in use belongs to a volume group (VG). All of the physical volumes in a volume group are divided into physical partitions (PPs) of the same size. For space allocation purposes, each physical volume is divided into five regions (outer_edge, inner_edge, outer_middle, inner_middle and center). The number of physical partitions in each region varies, depending on the total capacity of the disk drive.

Within each volume group, one or more logical volumes (LVs) are defined. Logical volumes are groups of information located on physical volumes. Data on logical volumes appears to be contiguous to the user but can be discontiguous on the physical volume. This allows file systems, paging space, and other logical volumes to be resized or relocated, to span multiple physical volumes, and to have their contents replicated for greater flexibility and availability in the storage of data.

Each logical volume consists of one or more logical partitions (LPs). Each logical partition corresponds to at least one physical partition. If mirroring is specified for the logical volume, additional physical partitions are allocated to store the additional copies of each logical partition. Although the logical partitions are numbered consecutively, the underlying physical partitions are not necessarily consecutive or contiguous.

Logical volumes can serve a number of system purposes, such as paging, but each logical volume serves a single purpose only. Many logical volumes contain a single journaled file system (JFS or JFS2). Each JFS consists of a pool of page-size (4 KB) blocks. When data is to be written to a file, one or more additional blocks are allocated to that file. These blocks might not be contiguous with one another or with other blocks previously allocated to the file. A given file system can be defined as having a fragment size of less than 4 KB (512 bytes, 1 KB, 2 KB).

The system has one volume group called `rootvg` that consists of a base set of logical volumes required to start the system. Any other physical volumes you have connected to the system can be added to a volume group using the **`extendvg`** command. You can add the physical volume either to the `rootvg` volume group or to another volume group by using the **`mkvg`** command. Logical volumes can be tailored using the commands, the menu-driven System Management Interface Tool (SMIT) interface, or Web-based System Manager.

Logical volume concepts

The basic logical storage concepts are as follows: Physical Volumes, Volume Groups, Physical Partitions, Logical Volumes, and Logical Partitions.

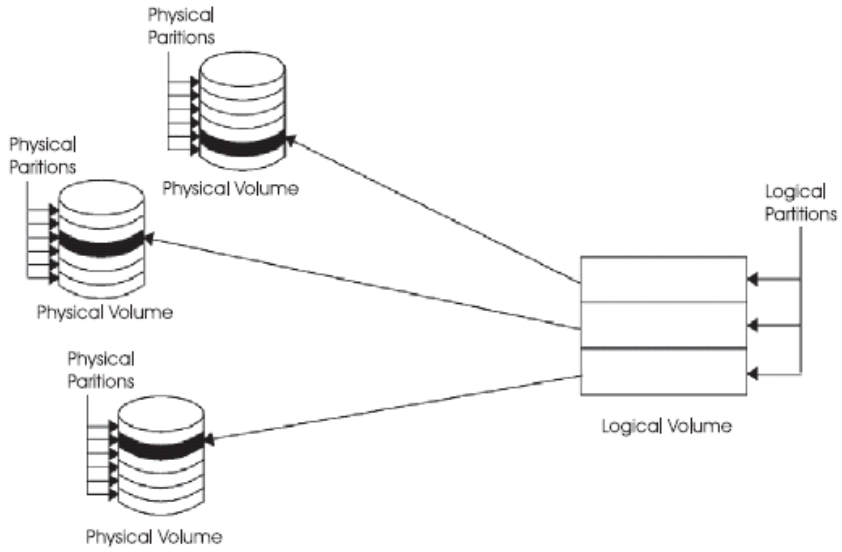


Figure 1 Basic logical storage concepts

Physical volumes

Each disk must be designated as a physical volume and be put into an available state before it can be assigned to a volume group. A physical volume has certain configuration and identification information written on it which includes a physical volume identifier (PVID) that is unique to the system.

In AIX 5.2 and later versions, the LVM can make use of the additional space that can be added to a logical unit number (LUN), by adding physical partitions to the physical volume associated with the LUN. For more information, see “LUN expansion” on page 203.

Volume groups

A normal volume group is a collection of 1 to 32 physical volumes of varying sizes and types. A big volume group can have from 1 to 128 physical volumes. A scalable volume group can have up to 1024 physical volumes. A physical volume can belong to only one volume group per system; there can be up to 255 active volume groups.

When a physical volume is assigned to a volume group, the physical blocks on the disk are organized into physical partitions of a size you specify when you create the volume group.

One volume group (the root volume group, called rootvg) is configured automatically during installation time and contains the base set of logical volumes required to start the system. The rootvg includes paging space, the journal log, boot data, and dump storage, each in its own separate logical volume. The rootvg has attributes that differ from user-defined volume groups. For example, the rootvg cannot be imported or exported. When performing a command or procedure on the rootvg, you must be familiar with its unique characteristics.

You create a volume group with the **mkvg** command. You add a physical volume to a volume group with the **extendvg** command, make use of the changed size of a physical volume with the **chvg** command, and remove a physical volume from a volume group with the **reducevg** command. Some of the other commands that you use on volume groups include: list (lsvg), remove (exportvg), install (importvg), reorganize (reorgvg), synchronize (syncvg), make available for use (varyonvg), and make unavailable for use (varyoffvg).

You can move data from one physical volume to other physical volumes in the same volume group with the **migratepv** command. This command allows you to free a physical volume so it can be removed from the volume group. For example, you could move data from a physical volume that is to be migrated.

A volume group that is created with smaller physical and logical volume limits can be converted to a format which can hold more physical volumes and more logical volumes. This operation requires enough free partitions on every physical volume in the volume group for the volume group descriptor area (VGDA) expansion. The number of free partitions required depends on the size of the current VGDA and the physical partition size. Because the VGDA resides on the edge of the disk and it requires contiguous space, the free partitions are required on the edge of the disk. If those partitions are allocated for a user's use, they are migrated to other free partitions on the same disk. The rest of the physical partitions are renumbered to reflect the loss of the partitions for VGDA usage. This renumbering changes the mappings of the logical to physical partitions in all the physical volumes of this volume group. If you have saved the mappings of the logical volumes for a potential recovery operation, generate the maps again after the completion of the conversion operation. Also, if the backup of the volume group is taken with map option and you plan to restore using those maps, the restore operation might fail because the partition number might no longer exist (due to reduction). It is recommended that backup is taken before the conversion and right after the conversion if the map option is used. Because the VGDA space has been increased substantially, every VGDA update operation (creating a logical volume, changing a logical volume, adding a physical volume, and so on) might take considerably longer to run.

Physical partitions

When you add a physical volume to a volume group, the physical volume is partitioned into contiguous, equal-sized units of space called physical partitions. A physical partition is the smallest unit of storage space allocation and is a contiguous space on a physical volume.

Physical volumes inherit the volume group's physical partition size, which you can set only when you create the volume group (for example, using the **mkvg -s** command).

Logical volumes

After you create a volume group, you can create logical volumes within that volume group. A logical volume appears to users and applications as a single, contiguous, extensible disk volume. You can create additional logical volumes with the **mklv** command. This command allows you to specify the name of the logical volume and define its characteristics, including the number and location of logical partitions to allocate for it.

After you create a logical volume, you can change its name and characteristics with the **chlv** command, and you can increase the number of logical partitions allocated to it with the **extendlv** command. The default maximum size for a logical volume at creation is 512 logical partitions, unless specified to be larger. The **chlv** command is used to override this limitation.

Logical volumes can also be copied with the **cplv** command, listed with the **lslv** command, removed with the **rmlv** command, and have the number of copies they maintain increased or decreased with the **mklvcopy** and the **rmlvcopy** commands. Logical Volumes can also be relocated when the volume group is reorganized.

The system allows you to define up to 255 logical volumes per standard volume group (511 for a big volume group and 4095 for a scalable volume group), but the actual number you can define depends on the total amount of physical storage defined for that volume group and the size of the logical volumes you define.

Logical partitions

When you create a logical volume, you specify the number of logical partitions for the logical volume. A logical partition is one, two, or three physical partitions, depending on the number of instances of your data you want maintained. Specifying one instance means there is only one copy of the logical volume (the default). In this case, there is a direct mapping of one logical partition to one physical partition. Each instance, including the first, is termed a copy. Where physical partitions are located (that is, how near each other physically) is determined by options you specify when you create the logical volume.

File systems

The logical volume defines allocation of disk space down to the physical-partition level. Finer levels of data management are accomplished by higher-level software components such as the Virtual Memory Manager or the file system. Therefore, the final step in is the creation of file systems. You can create one file system per logical volume. To create a file system, use the **crfs** command.

Limitations for logical storage management

Table 3 shows the limitations for logical storage management. Although the default maximum number of physical volumes per volume group is 32 (128 for a big volume group, 1024 for a scalable volume group), you can set the maximum for user-defined volume groups when you use the **mkvg** command. For the **rootvg** command, however, this variable is automatically set to the maximum by the system during the installation.

Table 3 Logical storage management limits

Category	Limit
Volume group	255 per system

Table 3 Logical storage management limits

Category	Limit
Physical volume	(MAXPVS/volume group factor) per volume group. MAXPVS is 32 for a standard volume group, 128 for a big volume group, and 1024 for a scalable volume group.
Physical partition	Normal and Big Volume groups: (1016 x volume group factor) per physical volume up to 1024 MB each in size. Scalable volume groups: 2097152 partitions up to 128 GB in size. There is no volume group factor for scalable volume groups.
Logical volume	MAXLVS per volume group, which is 255 for a standard volume group, 511 for a big volume group, and 4095 for a scalable volume group.

If you previously created a volume group before the 1016 physical partitions per physical volume restriction was enforced, stale partitions (no longer containing the most current data) in the volume group are not correctly tracked unless you convert the volume group to a supported state. You can convert the volume group with the **`chvg -t`** command. A suitable factor value is chosen by default to accommodate the largest disk in the volume group.

For example, if you created a volume group with a 9 GB disk and 4 MB partition size, this volume group will have approximately 2250 partitions. Using a conversion factor of 3 (1016 * 3 = 3048) allows all 2250 partitions to be tracked correctly. Converting a standard or big volume group with a higher factor enables inclusion of a larger disk of partitions up to the 1016* factor. You can also specify a higher factor when you create the volume group in order to accommodate a larger disk with a small partition size.

These operations reduce the total number of disks that you can add to a volume group. The new maximum number of disks you can add would be a MAXPVS/factor. For example, for a regular volume group, a factor of 2 decreases the maximum number of disks in the volume group to 16 (32/2). For a big volume group, a factor of 2 decreases the maximum number of disks in the volume group to 64 (128/2). For a scalable volume group, a factor of 2 decreases the maximum number of disks in the volume group to 512 (1024/2).

Logical Volume Manager components

The set of operating system commands, library subroutines, and other tools that allow you to establish and control logical volume storage is called the Logical Volume Manager (LVM). The LVM controls disk resources by mapping data between a more simple and flexible logical view of storage space and the actual physical disks. The LVM does this using a layer of device-driver code that runs above traditional disk device drivers. The LVM consists of the logical volume device driver (LVDD) and the LVM subroutine interface library. The logical volume device driver (LVDD) is a pseudo-device driver that manages and processes all I/O. It translates logical addresses into physical addresses and sends I/O requests to specific device drivers. LVM subroutine interface library contains routines that are used by the system management commands to perform system management tasks for the logical and physical volumes of a system.

Useful AIX functions and utilities

Table 4 lists AIX functions and utilities you can use to define and manage Dell EMC devices. The use of these functions and utilities is optional; they are listed for reference only.

Table 4 AIX functions and utilities (page 1 of 2)

Function or utility	Definition
`SMIT`	Menu-driven utility for managing the device configuration of the AIX server. The System Management Interface Tool (`SMIT`) is available in a windows version (**`xinit`** command) or an ASCII interface. SMIT can be used for a number of standard AIX functions, such as: • Installing and maintaining software • Installing and managing devices • Managing disk devices using the LVM • Problem determination
`cfgmgr -v`	Configuration manager with verbose parameter. This is an inline hardware resource utility that *walks* the bus(ses) and examines all adapters and devices connected to the AIX server. It is frequently used in conjunction with the **`lsdev`** and **`lscfg`** commands.
`lsdev -Cs scsi`	Lists available SCSI device(s).
`lsdev -Cs fcscsi`	Use with the IBM driver to list available Fibre Channel devices.
`lsdev -Cs fcs`	Use with the IBM driver to list all Fibre Channel devices in the system.
`lsdev -Cc disk`	Lists all disk devices.
`lsdev -C | more`	Lists all devices and directs the output to the screen one page at a time.
`lscfg | grep scsi`	Lists the configuration in PCI-based machines. It shows all SCSI channel host controller cards available or defined in the system.
`lsdev -Ct efscsi`	Use with the IBM driver to list all Fibre Channel host controller protocol devices.
`lsdev -Ct df1000f7`	Use with the IBM driver to list all Fibre Channel host controller cards available or defined in the system.
`lsattr -E -l scsi#`	Lists the attributes (addresses) of the SCSI number requested on the Fast-Wide Differential host controller card.
`lsattr -El fcs#`	Use with the IBM driver to list the attributes (addresses) of the `fcs` number requested on the Fibre Channel host controller card.
`lsattr -El fscsi#`	Use with the IBM driver to list the attributes (addresses) of the `fscsi` number requested on the Fibre Channel host controller card.
`lsvg -o`	Displays all volume groups that are varied on (activated).
`lsvg -l vgname`	Displays logical volumes, physical partitions, and the state of the logical volume within the specified volume group.
`mkdev`	Configures a disk device, making sure it is available as a physical volume.
`mkvg`	Groups one or more physical disks into a volume group.
`varyonvg`	Makes the specified volume group available to the host.
`varyoffvg`	Makes the specified volumes group unavailable to the host.
`crfs`	Creates a new filesystem.

Table 4 AIX functions and utilities (page 2 of 2)

Function or utility	Definition
`mount`	Mounts a filesystem.
`errpt -aN hdisk`*X* *	Generates an error report that lists all errors associated with the `hdisk`*X*.
`odmget` *XxXx*	Allows you to see all hardware and devices configured, defined, or available in the system. Predefined attributes are supported devices in the `/dev` library. Customized attributes are all configured, defined, or available devices. All customized attributes are located in the `/etc/objrepos` library. where *XxXx* is one of the following Object Database Manager (ODM) attributes: • `CuAt` — ODM Customize attributes • `PdAt` — Predefined attributes • `CuDv` — Customize devices • `PdDV` — Predefined attributes
	IMPORTANT Inappropriate use of the following rmdev commands can result in a system problem and/or loss of data. Take extra caution when using this command to ensure that the devices and/or controllers you are working with are not in use and are the correct addresses.
`rmdev -l scsi`*X* `-d`	Removes a SCSI controller (as specified by the *X* in `scsi`*X*) from the ODM.
`rmdev -l hdisk`*X*	Removes the specified disk device.
`rmdev -l hdisk`*X* `-d`	Removes the specified disk device (`hdisk`*X*) and deletes it from the ODM.
`rmdev -l fscsi`*X* `-d`	Removes the IBM Fibre Channel protocol driver instance from the ODM.
`rmdev -l fcs`*X* `-d`	Use with the IBM driver to remove the Fibre Channel controller *X* from the ODM.

Operating system updates

The operating system package is divided into filesets, where each fileset contains a group of logically related customer deliverable files. Each fileset can be individually installed and updated.

Revisions to filesets are tracked using the version, release, maintenance, and fix (VRMF) levels. By convention, each time an AIX fileset update is applied, the fix level is adjusted. Each time an AIX maintenance level is applied, the modification level is adjusted, and the fix level is reset to zero. The initial installation of an AIX version, for example, AIX 5.2, is called a base installation. The operating system provides updates to its features and functionality, which might be packaged as a maintenance level, a program temporary fix (PTF), or a recommended maintenance package.

Maintenance levels

A maintenance level provides new functionality that is intended to upgrade the release. The maintenance part of the VRMF is updated in a maintenance level. For example, the first maintenance level for AIX 5.2 would be 5.2.1.0; the second would be 5.2.2.0, and so forth.

PTFs

Between releases, you might receive PTFs to correct or prevent a particular problem. A particular installation might need some, all, or even none of the available PTFs.

Recommended maintenance packages

A recommended maintenance package is a set of PTFs between maintenance levels that have been extensively tested together and are recommended for preventive maintenance.

Determining what is installed

To determine the maintenance level installed on a particular system, type:

```
oslevel
```

To determine which filesets need updating for the system to reach a specific maintenance level (in this example, `5.1.1.0`), use the following command:

```
oslevel -l 5.1.1.0
```

To determine if a recommended maintenance package is installed (in this example, `5100-02`), use the following command:

```
oslevel -r 5100-02
```

To determine which filesets need updating for the system to reach the `5100-02` level, use the following command:

```
oslevel -rl 5100-02
```

To determine the level of a particular fileset (in this example, `bos.mp`), use the following command:

```
lslpp -L bos.mp
```

OS requirements for support of more than 2 TB LUNs

AIX 5L v5.1 and higher provides support for LUNs larger than 2 TB with the 64-bit kernel. Dell EMC supports this feature for Dell EMC VMAX™, Dell EMC Symmetrix™, EMC VNX™, and Dell EMC CLARiiON™ systems.

Minimum OS release levels:

- AIX 5.1 ML-08 (APAR IY70781) with APAR IY66914
- AIX 5.2 ML-05 (APAR IY66260)
- AIX 5.3 ML-02 (APAR IY69190)

The following output example output from an **inq** command shows a capacity of a LUN greater than 2 TB:

```
./inq.aix64
Inquiry utility, Version V7.3-487 (Rev 0.0)(SIL Version V5.3.0.0 (Edit Level 487)
Copyright (C) by EMC Corporation, all rights reserved.
For help type inq -h.
...........
-------------------------------------------------------------------------------
DEVICE              :VEND  :PROD     :REV   :SER NUM      :CAP(kb)
-------------------------------------------------------------------------------
/dev/rhdiskpower0   :DGC   :RAID 5   :0219  :3D000077     :2147483648
/dev/rhdiskpower1   :EMC   :SYMMETRIX:5771  :1201B640     :2147483648
/dev/rhdisk2        :DGC   :RAID 5   :0219  :3D000077     :2147483648
/dev/rhdisk3        :EMC   :SYMMETRIX:5771  :1201B640     :2147483648
```

Installation methods

This section describes three methods of installation:

- "New installation" on page 29

  This method assumes that you have an existing AIX installation and are migrating that installation to the VNX series or CLARiiON or Symmetrix boot device.

- "Migration installation" on page 30

  This method assumes that this a new installation of AIX directly to the external VNX series or CLARiiON or Symmetrix boot device.

- "Cloned installation" on page 30

  This method assumes that the internal installation will be cloned to the external device using IBM's alt_disk_install package, leaving two installations to alternate between. The alternate disk installation is the preferred method since it leaves a backup installation of AIX on the host.

Prior to configuring a system for booting from the VNX series or CLARiiON or Symmetrix array, microcode updates are required in two areas:

- The adapter microcode is placed on the adapter that is to become your bootable adapter. There are different microcode levels based on the adapter that you will use for booting.

  The Dell EMC Simple Support Matrices lists the HBAs currently supported for Fibre Channel boot, as well as the required HBA firmware.

- The system/service processor combined microcode is based on the model number of the host that will be used for Fibre Channel boot (for example, 7017-S85).

All firmware updates can be obtained from this website:

`http://techsupport.services.ibm.com/server/mdownload/download.html`

This link also provides complete instructions on updating the firmware and verifying that the process was successful. Many AIX hosts now come with supported revisions of firmware already installed on both the hosts and the adapters. If your hosts already meet minimum requirements, any steps for installing host or adapter firmware can be safely skipped.

New installation

1. Identify the host to be used for Fibre Channel boot.
2. Identify the type of host adapter to be used for Fibre Channel boot and install it.
3. Update the adapter microcode as described under "Host with existing adapters" on page 32.
4. Update the system/service processor microcode as described under "Updating system/service processor-combined microcode" on page 33.
5. Perform the installation to the external boot device as described under "New installation" on page 36.
6. Install the Dell EMC Object Data Manager Support Package as described under "Installing the Dell EMC ODM support package" on page 40.

Migration installation

1. Identify the host to be used for Fibre Channel boot.
2. Identify the type of host adapter to be used for Fibre Channel boot and install it.
3. Update the adapter microcode as described under “Host with existing adapters” on page 32.
4. Update the system/service processor microcode as described under “Updating system/service processor-combined microcode” on page 33.
5. Perform a migration install to the AIX installation currently installed to your internal disk.
6. Use migration procedures to migrate from your internal hdisk device to the external Symmetrix device as described under “Migration installation” on page 38.
7. Install the Support Package as described under “Installing the Dell EMC ODM support package” on page 40.

Cloned installation

1. Identify the host to be used for Fibre Channel boot.
2. Identify the type of host adapter to be used for Fibre Channel boot and install it.
3. Update the adapter microcode as described under “Host with existing adapters” on page 32.
4. Update the system/service processor microcode as described under “Updating system/service processor-combined microcode” on page 33.
5. Perform the clone from the current boot volume to the external disk as described under “Cloned installation using alt_disk_install” on page 38.
6. Install the ODM Support Package as described under “Installing the Dell EMC ODM support package” on page 40.

Preparing your system for booting from the Dell EMC storage array

This section contains information to prepare your system for booting from the Dell EMC storage array.

Prerequisites

- A supported AIX server system with AIX installed to the internal disk drive.
- A minimum of one supported host adapter, to be used as the boot adapter.
- An Internet connection or method for getting the microcode update files onto the host.
- The installation CD for your revision of AIX or the IBM APAR containing the drivers for the HBA you are using.

Note: To get APARs, go to IBM's fix central and use the search feature:

http://www-912.ibm.com/eserver/support/fixes/fixcentral/fixdetails?fixid=SI22191

Recommendations

- If the host does not have the adapter(s) installed, install only the adapter to be used for booting at this time.
- If the host already has multiple adapters installed, to avoid confusion, you may want to remove all Fibre Channel adapters except the one intended to be used as the boot adapter.

Adding the adapters to your host

1. Identify the feature code of the adapter you will use for Fibre Channel boot.
2. Insert the adapter into the host. If the system is not hot-pluggable, you will have to power down the host to insert the adapter. Once the adapter is installed, you must install the drivers for it. If the adapter has already been installed in the system and you know the drivers are loaded, you may skip to the section "Host with existing adapters" on page 32.
3. Insert the installation CD (or point to the location of the APAR files) and run `smitty install`.
4. Select **Install and Update Software**.
5. Select **Install and Update from LATEST Available Software**.
6. Select your installation media.
7. Search for the apropriate fileset using the menu. Once you find it, mark it and complete the install.
8. For complete functionality, a reboot is recommended.
9. After a successful reboot, the operating system detects the fibre adapter(s). Use the **lsdev** command as follows to display the fibre adapters. Note the name of the adapter on your system.

```
# lsdev -Cc adapter | grep fcs
fcs0    Available 10-68    FC Adapter
fcs1    Available 10-78    FC Adapter
```

10. Proceed to the next section, "Host with existing adapters" on page 32, and begin with step 2.

Host with existing adapters

1. Enter the command **lsdev -Cc adapter | grep fcs**. If the command returns no results, refer to Adding the Adapters to Your Host. In our example, we will be using two installed adapters. FCS0 will be used as the boot adapter. Note the name of the adapters.

```
# lsdev -Cc adapter | grep fcs
fcs0    Available 10-68    FC Adapter
fcs1    Available 10-78    FC Adapter
```

Note: If there are devices allocated to the adapter with volume groups varied on, they must be varied off before continuing.

2. Go to:

http://techsupport.services.ibm.com/server/mdownload/download.html

3. Click the **Adapter Microcode** link, and locate the feature code for the adapter you will be using. You can first click on the description file, which contains instructions for *Determining Adapter Microcode Levels and Updating the Adapter's Microcode*. You should either print the file or make sure you have access to it throughout the update process.
4. After you have familiarized yourself with the update procedure, download the adapter microcode.
5. Use the description file containing the instructions to perform the update.
6. After the update is complete, look at the configuration for your adapter and verify that the levels meet the required minimums in the Dell EMC Simple Support Matrices.

Updating system/service processor-combined microcode

This section contains information to update the system/service processor-combined microcode.

Prerequisites

- A supported AIX server system with AIX installed to the internal disk drive.
- Host model number (for example, 7017-S85).
- An Internet connection or method for getting the microcode update files onto the host.

Procedure

1. Go to:

   http://www.rs6000.ibm.com/support/micro/download.html

2. Under system/service processor-combined microcode, locate the model number of the host you will be using. You can first click on the description file, which contains instructions for downloading microcode to the system/service processor. You should either print the file or make sure you have access to it throughout the update process.
3. After you have familiarized yourself with the update procedure, download the system/service processor combined microcode for your host.
4. Use the description file containing the instructions to perform the update.
5. A reboot is required for this microcode update to take full effect.

Connecting to the VNX series or CLARiiON system

This section contains information to connect to the VNX series or CLARiiON system.

Prerequisites

- A supported AIX server system with AIX installed to the internal disk drive with SAN boot approved host microcode installed and active.

  Note: See the Dell EMC Simple Support Matrices for a list of supported hosts.

- An external boot-supported HBA.
- Appropriate operating system drivers for your HBA.

Adding physical connections

1. Start by adding all desired physical connections between the host and the array.
2. For switched fabric, zones must be created to allow the host's HBA(s) to log into the array's SP ports. If a host is to have the visibility into multiple SP ports, make sure that the zones include all of the expected members.
3. Using whatever management utility appropriate for your topology and equipment, verify that all host and array components have logged in and all zones have been created allowing the host's visibility into the array. To verify this visibility:
   a. Log in to Unisphere/Navisphere Manager and right-click the **Array** component under the **Storage** tab.
   b. Select **Connectivity Status**. If this is the first time connecting the host to the array, you should see the WWN for each HBA showing up as Fibre Logged In > Yes and Registered > No. (If this is a host that had existing connections, this does not apply.)
4. If you see all the hosts HBA(s) logged in, your physical connections are complete. If all of the adapters have not logged in, you must figure out where the breakdown is and address it before continuing.

Completing your configuration

If the host agent is loaded, some of this process may take place automatically. If the host agent is not loaded, you must register your adapters manually.

1. If your adapters have not been registered before, register them as follows:
   a. Return to the **Connectivity Status** screen.
   b. Right-click the **Array** component under the **Storage** tab, and then select **Connectivity Status**.
2. Select each adapter and, in turn, click **Register** at the bottom of the window.
3. In the following window, select **AIX hosts Initiator Type=CLARiiON Open**, **Failover Mode=3**, **ArrayCommPath=disabled**. These settings are the appropriate settings for Dell EMC PowerPath™ configurations. Enter any other information such as hostname and IP address.

Note: For a more complete explanation of these settings and how to set them, refer to *Setting Storage System Properties* in the *EMC Storage-System Host Utilities for AIX Administrator's Guide*.

4. Run configuration manager against each adapter.

```
# cfgmgr -vl fcs0
Time: 1 LEDS: 0x539
Number of running methods: 0
----------------
attempting to configure device 'hdisk1'
Time: 1 LEDS: 0x626
invoking /usr/lib/methods/cfgscsidisk -l hdisk1
Number of running methods: 1
```

5. For each initiator you should be able to use the **lsdev** command as follows. Verify your connections to the VNX series or CLARiiON SP.

```
# lsdev -Cc array
sp0 Available  CLARiiON Storage Processor
sp1 Available  CLARiiON Storage Processor
```

6. VNX series or CLARiiON Array CommPath behavior results in devices appearing visible to a host when no LUNs are bound. There is one device visible per initiator (path). Each device has an ID of drivetype unknown. For AIX, as mentioned above, the arraycommpath setting should be disabled. In the event that you do see LUNZ devices, you must `rmdev` these in order to properly utilize your storage groups once they are configured.

```
# lsdev -Cc disk
hdisk1 Available 30-68-00-10,0 16 Bit SCSI Disk Drive
hdisk0 Available 30-68-00-9,0  16 Bit LVD SCSI Disk Drive
hdisk2 Available 50-58-01      FC CLArray LUNZ Disk
hdisk3 Available 80-60-01      FC CLArray LUNZ Disk
```

7. You must now go through the standard VNX series or CLARiiON practice:

   a. Create your RAID group(s).

   b. Create LUNs within your RAID group(s).

   c. Create storage group(s).

   d. Assign your LUNs into your storage group(s).

   e. Add your host to your storage group(s) for access to your LUNs.

   For detailed instructions on each of these, refer to the *EMC ControlCenter Navisphere Manager Administrator's Guide*.

8. After you complete step 7 a through e, be sure to use the **rmdev** command to remove any LUNZ devices prior to configuring you real LUNs on the host.

9. When all of these steps are complete, you can create your AIX hdisk devices from the LUNs that are presented to the host by the VNX series or CLARiiON. These steps are outlined in "Installing to the Symmetrix or VNX series or CLARiiON disk" on page 36.

Installing to the Symmetrix or VNX series or CLARiiON disk

Even in an environment where you plan on doing a new installation to the external boot device, you may have devices capable of being used as a boot device already allocated to the host. For this reason, and since SANs by nature allow access to a large number of devices, identifying the hdisk to install to can be difficult. Carefully review the following recommendations so you can easily discover the lun_id to hdisk correlation.

Prerequisites

- A supported AIX server system with AIX installed to the internal disk drive.
- A Fibre Channel boot-supported HBA with the appropriate firmware revision.
- Physical connections in place to the array or switch, depending on which topology you will be using.

Recommendations

- Zone the switch or disk array such that the machine being installed can only discover the disk(s) to be installed to. After the installation has completed, you can re-open the zoning so the machine can discover all necessary devices.
- Assign PVID (physical volume identifiers) to all disks from an already-installed AIX system that can access the disks. To do this, use the command **`chdev -l hdiskX -a pv=yes`**, where *`X`* is the appropriate number. Create a table mapping PVIDs to physical disks. The PVIDs will be visible from the install menus by selecting the alternate attributes option when selecting the disk to install to.
- Use ODM commands to locate the hdisk to install to. From the main install menu, select **Start Maintenance Mode for System Recovery**, **Access Advanced Maintenance Functions**, and enter the **Limited Function Maintenance Shell**. At the prompt, you can use the command **`odmget -q"attribute=lun_id AND value=0xNN..N" CuAt`**, where *`0xNN..N`* is the lun_id you are looking for. This command will print out the ODM stanzas for the hdisks that have this lun_id. Enter **exit** to return to the installation menus.

  As of this writing, AIX does not support multiple paths to the boot device without multi-pathing software. Attach the external disk and the adapter into the SAN Switch. Verify that the adapters have logged in to the switch and the World Wide Names match the new adapters. These may be labeled on the back of the physical adapter, or can be found with the **`lscfg -pv | pg`** command and search for the `fcs#` stanza of the adapter you are using. The Network Address field in the **`lscfg`** output is equal to the World Wide Name. Once you have attached the disks, run `cfgmgr` on the AIX host to add the newly discovered disks into the servers ODM. It is a good idea to only attach one physical connection at this time, even if you are later going to use a multipathing solution (that is, PowerPath) due to problems that could occur during installation.

New installation

1. If a boot device has already been assigned and discovered by the host via the internal installation of AIX, skip to Step 3.

2. After the boot device has been assigned to the host, you must discover the boot device. To do so, use the command **cfgmgr –vl fcs*X***, where *X* is the number of the adapter that will be used for Fibre Channel boot. As the following output shows, unless you already have the EMC ODM Support Package installed, the new external devices will appear as Other FC SCSI Disk Drive. Only the relevant portions of the output are shown.

```
# lsdev -Cc disk
hdisk0 Available 30-68-00-10,0 16 Bit LVD SCSI Disk Drive

# cfgmgr -vl fcs0
Time: 1 LEDS: 0x539
Number of running methods: 0
---------------
attempting to configure device 'hdisk1'
Time: 1 LEDS: 0x626
invoking /usr/lib/methods/cfgscsidisk -l hdisk1
Number of running methods: 1

# lsdev -Cc disk
hdisk0 Available 30-68-00-10,0 16 Bit LVD SCSI Disk Drive
hdisk1 Available 10-58-01      Other FC SCSI Disk Drive
```

3. The easiest way to track your external boot device is to assign a PVID to the intended installation disk. This is done using the command **chdev -l hdiskX -a pv=yes**.

```
# lspv
hdisk0          000c53cdd955af6e    rootvg
hdisk1          none                None

# chdev -l hdisk1 -a pv=yes
hdisk1 changed

# lspv
hdisk0          000c53cdd955af6e    rootvg
hdisk1          000c53cde980ba9d    None
```

4. Write down the PVID; you will then use this to identify your installation disk.
5. Place the AIX installation media into the host and shut down. After the shutdown, you can remove the internal disk if you want.
6. Boot from the installation media. If the installation media was not a part of your bootlist, you can select it by entering the SMS menus and selecting your installation device as a boot device. Typically, the CD-ROM is a part of your bootlist.
7. Select the display to be used as a terminal.
8. Select the language to be used.
9. Select option 2, **Change/Show Installation Settings and Install**.
10. Select option 1, **System Settings**, and ensure that **New and Complete Overwrite** is selected.
11. If you have more than one disk allocated to the adapter, to verify your installation disk you can select option 77, **Display More Disk Information**, to see the PVID of the hdisks. Depending on the method you used to identify the installation disk, as you continue selecting 77, additional information such as WWPN, SCSI ID, and LUN ID will appear. Also, on the initial disk selection screen, verify that **Yes** appears under the field marked bootable for the hdisk you will install to. Once you have identified the correct hdisk device(s), select it, and then continue. After the installation is complete, the host will automatically reboot from your external boot device.

Migration installation

1. Unlike the new installation procedure, because you will be migrating from an internal hdisk there are several ways to update your existing version of AIX. Using your preferred method, update or migrate your current version of AIX. The update must be made prior to attempting to boot the operating system from the external device.
2. After the migration or update is complete and the system rebooted, begin the migration to external disk by adding the external boot device to the root volume group. For the remaining commands, hdisk0 is the internal boot device we are migrating from and hdisk1 is the external boot device we will migrate to.

   ```
   # extendvg -f rootvg hdisk1
   ```

3. To determine if the boot image is on the disk you want to migrate, run the **lslv –l bootlv** command:

   ```
   # lslv -l hd5
   hd5:N/A
   PV                COPIES        IN BAND       DISTRIBUTION
   hdisk0            001:000:000   100%          001:000:000:000:000
   ```

4. Migrate the boot logical volume to the external boot device:

   ```
   # migratepv –l hd5 hdisk0 hdisk1
   ```

5. After you have moved the boot image, clear the boot record on the original disk to prevent possible problems should the host accidentally attempt to boot from that disk:

   ```
   # mkboot –cd /dev/hdisk0
   ```

6. Update the boot image on your external boot device:

   ```
   # bosboot –ad /dev/hdisk1
   ```

7. At this point, you can migrate the rest of the of the contents of the source physical disk to the external boot device:

   ```
   # migratepv hdisk0 hdisk1
   ```

8. Use the **reducevg** command to remove the original disk device from the root volume group:

   ```
   # reducevg –d rootvg hdisk0
   ```

9. Use the **bootlist** command to change the bootlist to reflect the new boot device:

   ```
   # bootlist –m normal hdisk1
   ```

10. Reboot the host:

    ```
    # shutdown –Fr now
    ```

Cloned installation using alt_disk_install

1. If you have not already installed the alt_disk_install LPP, begin by selecting that fileset and installing it. It can be located on the AIX installation CD.

   ```
   # installp -a -d/dev/cd0 bos.alt_disk_install
   ```

2. After the fileset has been successfully installed, rootvg can now be cloned to the external disk as opposed to moved, leaving the internal original copy installed to the internal disk. You should notice that this process automatically updates the bootlist for you to the device that you cloned to.

```
# alt_disk_install -C hdisk1
Calling mkszfile to create new /image.data file.
Checking disk sizes.
Creating cloned rootvg volume group and associated logical volumes.
Creating logical volume alt_hd5
Creating logical volume alt_hd8
Creating logical volume alt_hd6
Creating logical volume alt_hd4
Creating logical volume alt_hd1
Creating logical volume alt_hd3
Creating logical volume alt_hd2
Creating logical volume alt_hd9var
Creating /alt_inst/ file system.
Creating /alt_inst/home file system.
Creating /alt_inst/tmp file system.
Creating /alt_inst/usr file system.
Creating /alt_inst/var file system.
Generating a list of files
for backup and restore into the alternate file system...
Backing-up the rootvg files and restoring them to the
alternate file system...
Modifying ODM on cloned disk.
Building boot image on cloned disk.
forced unmount of /alt_inst/var
forced unmount of /alt_inst/usr
forced unmount of /alt_inst/tmp
forced unmount of /alt_inst/home
forced unmount of /alt_inst
forced unmount of /alt_inst
Changing logical volume names in volume group descriptor area.
Fixing LV control blocks...
Fixing file system superblocks...
Bootlist is set to the boot disk: hdisk1
```

3. From this point forward, whichever hdisk you actively boot from will contain `rootvg`. Whenever you boot from your original installation, the alternate hdisk will contain a volume group called `altinst_rootvg`.

```
# lspv
hdisk0          0002583f112f6d71    rootvg
hdisk1          0002583f69d5822b    altinst_rootvg
```

4. However, after booting from the cloned installation, you will notice that the original installation will be called `old_rootvg`. Reboot the host now to use the external boot device, and booted from hdisk1. Updating the bootlist to use whichever device you want to boot from does switching between the two installations.

```
# shutdown -Fr now
# lspv
hdisk0          0002583f112f6d71    old_rootvg
hdisk1          0002583f69d5822b    rootvg
```

Installing the Dell EMC ODM support package

This section contains information to install the Dell EMC ODM support package.

The following filesets are distributed within the support package. Always refer to the Dell EMC Support Matrices for the most up-to-date supported solutions.

- Symmetrix AIX Support Software
  - Standard utility support
- Symmetrix FCP Support Software
  - IBM Fibre Channel driver support.
  - Supports PowerPath
- Symmetrix FCP MPIO Support Software
  - IBM default PCM MPIO support
- Symmetrix FCP Power MPIO Support Software
  - PowerPath custom PCM MPIO
  - Not supported at this time
- Symmetrix iSCSI Support Software
  - IBM iSCSI driver support
  - Supports PowerPath
- Symmetrix HA Concurrent Support
  - Requires HACMP CLVM
- CLARiiON AIX Support Software
  - Standard utility support for VNX series and CLARiiON
- CLARiiON FCP Support Software
  - IBM Fibre Channel driver support for VNX series and CLARiiON
  - Supports PowerPath for VNX series and CLARiiON
- CLARiiON FCP MPIO Support Software
  - IBM default PCM MPIO support for VNX series and CLARiiON
- EMC CLARiiON FCP PowerMPIO Support Software
  - PowerPath custom PCM MPIO support for VNX series and CLARiiON
  - Not supported at this time
- CLARiiON iSCSI Support Software
  - IBM iSCSI driver support for VNX series and CLARiiON
  - Supports PowerPath for VNX series and CLARiiON

- CLARiiON HA Concurrent Support
  - Requires HACMP CLVM for VNX series and CLARiiON
- Dell EMC Invista™ AIX Support Software
  - Standard utility support
- Invista FCP Support Software
  - IBM Fibre Channel driver support.
  - Supports PowerPath
- Dell EMC Celerra® AIX Support Software
  - Standard utility support
- Celerra FCP Support Software
  - IBM iSCSI driver support

Prerequisites

- A supported AIX server system with AIX installed to the internal or external boot device.
- An Internet connection or method for getting the ODM installation files onto the host.

ODM download and installation for Symmetrix, VNX series, and CLARiiON release 13 or higher

1. From the FTP server, `ftp.EMC.com`, in `/pub/elab/aix/ODM_DEFINITIONS`, download the most recent ODM fileset. If a more recent fileset is available, use it as a substitute.
2. In the `/tmp` directory, uncompress the fileset. Substitute the appropriate filename revision for the command to work properly:

   ```
   uncompress EMC.AIX.X.X.X.tar.Z
   ```

3. After you uncompress the fileset, you will be left with a file with the .tar extension. Untar the resulting file. Substitute the appropriate filename revision for the command to work properly:

   ```
   tar -xvf  EMC.AIX.X.X.X.tar
   ```

4. If the SMIT menu interface is preferred, invoke `smit installp` from the `/tmp` directory.

   Select **Install and Update from LATEST Available Software**.
5. Use the List function to select `/tmp` as the installation directory.
6. Use the List function to select Symmetrix AIX Support Software, CLARiiON Fibre Channel Support Software, or both depending on your configuration.
7. Press **Enter** *after* making all desired changes. You can use the default options.
8. Be sure to scroll all the way down to the bottom of the window to see the Installation Summary, and verify that the `SUCCESS` message is displayed.
9. Reboot the host for all changes to take effect.

Special considerations

This section contains information on the following for special consideration:

Note: Refer to Appendix Appendix A, "Migrating Considerations for AIX," when migrating AIX from one Dell EMC array to another.

PowerPath

On some storage systems, you can use a PowerPath hdiskpower device as a boot device—the device that contains the startup image. (Consult the Dell EMC Simple Support Matrices to find out if your storage system supports PowerPath boot devices.) Using a PowerPath hdiskpower device as a boot device provides load balancing and path failover for the boot device.

On AIX, booting from a PowerPath device is supported in SCSI and Fibre Channel environments that include specific versions of Dell EMC Enginuity™ software. Refer to the Dell EMC Simple Support Matrices for details. Contact your Dell EMC Customer Support representative for information about installing Enginuity software.

When you set up a Symmetrix, VNX series, or CLARiiON PowerPath boot device, consider the following:

- All path devices that make up the hdiskpower device must be valid AIX boot devices.
- The boot device should not be visible to any other host attached to the same storage system. If using a storage system device as a boot device in an HACMP environment (with or without PowerPath), other hosts should not be able to address the boot device.
- The host's boot list must contain all hdisks that compose the hdiskpower device being used as the boot device. Otherwise, the host may fail to boot if one or more paths is disabled while the machine tries to boot.
- At startup, the system searches for an AIX boot image in the boot list, which is a list of hdisks stored in the hardware's NVRAM. If the system fails to boot, you can change the boot list. Use one of the following methods:
  - Boot the system from an installation device (CDROM or tape) into Maintenance Mode. Select the option to access the root volume group, and then run the AIX **bootlist** command from the shell.
  - Enter the **System Management Services** menu when the system starts, and use the **Multiboot** menu options to change the boot list. This method is faster, but it is more difficult to determine the correspondence between devices listed in the menu and the storage-system device you want to add to or remove from the boot list.

Considerations for VNX series or CLARiiON storage

PowerPath can be used to enable multipathing and failover to an external boot device on a VNX series or CLARiiON system, but such a configuration has some functional limitations and extra configuration steps. The primary limitation to VNX series or CLARiiON boot is that the AIX host must have access to the active SP for the boot LUN in order to be able to boot without user intervention. If all paths to the boot LUN's default SP are failed or the LUN has trespassed to the alternate SP, the host will not boot until access to the LUN through the default SP is restored. Using a switched environment where all HBAs have access to the default SP of the boot LUN can reduce the effects of this restriction.

In a VNX series or CLARiiON environment, the bootlist for the AIX host should include all hdisks that correspond to active and passive paths of the boot LUN on all associated SPs. If PowerPath is not configured, these devices will show a PVID in the output of the **`lspv`** command. Passive hdisks will show a PVID of `None`. If PowerPath is configured, use the **`powermt display dev=n`** command to examine the boot device and determine which hdisks are active to the default SP.

Once the AIX host is up and running, PowerPath will enable it to survive path failures and trespasses of the boot device.

The pprootdev tool

The `pprootdev` tool changes AIX configuration rules and updates the boot image so the AIX Logical Volume Manager uses hdiskpower devices to vary on the `rootvg` the next time the system boots. The `pprootdev` tool cannot change the state of rootvg on a running system. It does, however, modify ODM data that other tools use to determine what devices `rootvg` is using. For this reason, some commands report information that may appear to be incorrect if they are run after pprootdev is run and before a system reboot.

Trespasses

When a trespass occurs in a VNX series or CLARiiON environment, a passive interface becomes the active interface. In this situation, `bosboot` will fail unless you transfer the `rootvg` PVID to the newly active interface. To do so, run the command **`emcpassive2active`**. After you have run the command, **`bosboot`** will succeed.

Run **`emcpassive2active`** whenever a trespass occurs.

Configuring a new PowerPath installation

If the system contains sufficient internal storage, install and configure the operating system on the internal device(s). Use the procedure described in "Configuring an existing PowerPath installation" on page 45, to clone the operating system image on the storage system. If there is insufficient internal storage, use the following procedure to install AIX directly onto a storage-system device and use PowerPath to manage multiple paths to the root volume group:

1. Start with a single connection to the storage subsystem. If you are using a switch, only one logical path should be configured.
2. Install AIX on a storage system device that is accessed via fibre adapter.
3. Install the current storage-system drivers.
4. Reboot the host.
5. Use **`rmdev -d`** to delete any hdisks in the `Defined` state.
6. Install PowerPath. Refer to Chapter 2 of the *PowerPath for UNIX Installation and Administration Guide*.
7. Connect remaining physical connections between the host and the storage system. If you are using a switch, update the zone definitions to the new configuration.
8. Configure an hdisk for each path. In Chapter 2 of the *PowerPath for UNIX Installation and Administration Guide*, refer to the section "Before You Install on an AIX Host".
9. Run the **`powermt config`** command.
10. Use the **`pprootdev`** command to set up multipathing to the root device. Enter: **`pprootdev on`**.
11. Use the **`bootlist`** command to add all active **`rootvg`** storage devices to the boot list.
12. Reboot the host.

Configuring an existing PowerPath installation

This section describes the process for converting a system that has AIX installed on an internal disk to boot from a logical device on a storage system. The steps in this process are:

- Transferring a complete copy of the operating system from an internal disk to a logical device on the storage system.
- Configuring PowerPath so the root volume group takes advantage of multipathing and failover capabilities.

EMC recommends that you use this procedure. In the event of a problem, you can revert operations to the host's internal disks. First, ensure that the AIX disk includes:

- A copy of the AIX `alt_disk_install` LPP. (The LPP is on the AIX installation CD.)
- The `rte` and `boot_images` filesets.

Then follow these steps:

1. Verify that all device connections to the storage system are established.
2. Verify that all hdisks are configured properly. (Refer to "Before You Install on an AIX Host" in the *PowerPath for UNIX Installation and Administration Guide*.)
3. Locate drives with adequate space. Enter:

   `bootinfo -s hdiskx`

   For this example, assume hdisks 132-134 are adequate with 8 GB total space.
4. Use the **`powermt display`** command to determine which hdiskpower device contains hdisk132 (the first hdisk identified in step 2) as well as all the path hdisks for that hdiskpower. Enter:

   `powermt display dev=hdisk132`

   The output looks similar to this:

```
Pseudo name=hdiskpower38
Symmetrix ID=000100006216
Logical device ID=006C
state=alive; policy=SymmOpt; priority=0; queued-IOs=0
==============================================================
---------------- Host --------------- - Stor - -- I/O Path - -- Stats ---
### HW       Path      I/O    Paths     Interf.   Mode      State Q-IOs Errors
==============================================================
0  fscsi0    hdisk1    32     FA        14bA      active    alive  0      0
1 fscsi1     hdisk2    23     FA        14bB      active    alive  0      0
1 fscsi1     hdisk3    14     FA        14bA      active    alive  0      0
0 fscsi0     hdisk4    1      FA        14bB      active    alive  0      0
```

   You see that hdiskpower38 contains hdisk132 and that the path hdisks for hdiskpower38 are hdisk132, hdisk223, hdisk314, and hdisk41.
5. Run the **`powermt config`** command.
6. Run the following command to ensure there are PVIDs on all devices:

```
lsdev -Ct power -c disk -F name | xargs -n1 chdev -apv=yes -l
```

7. Use the **`rmdev`** command to remove all PowerPath devices. PowerPath should remain installed, but all PowerPath devices must be deleted.

```
lsdev -Ct power -c disk -F name | xargs -n1 rmdev -l
```

8. Remove the powerpath0 device.

   ```
   rmdev -l powerpath0
   ```

9. Run the **`lsdev -Ct power`** command. The command output should not list any devices.

10. Determine which hdisk(s) on the storage system will receive a copy of the operating system. Run this command to create the copy on the storage system hdisk(s):

    ```
    alt_install_disk -C hdisk_list
    ```

11. Reboot the system.

12. Specify that all path hdisks identified in step 3 are included in the bootlist. Enter:

    ```
    bootlist -m normal hdisk132 hdisk223 hdisk314 hdisk41
    ```

    If you are booting from a VNX series or CLARiiON storage system, the bootlist for the AIX host should include all hdisks that correspond to active and passive paths of the boot LUN on all associated SPs.

13. Run the **pprootdev on** command.

14. Reboot the system.

VNX series or CLARiiON MPIO devices

Unlike a PowerPath device, a VNX series or CLARiiON MPIO device does not auto trespass back to its default owner when the path is restored. The LUN is owned by its current SP until the active path is broken, which then causes it to trespass to the other SP.

For example, a LUN whose default owner is SPA would trespass to SPB when its path to SPA is broken. This breakage in path can be caused by many factors including SP reboots, disconnection of cables, or HBA or SAN switch failure. When the path is restored, the LUN continues to be owned by SPB unless a manual trespass on the LUN is made through Dell EMC Unisphere™, Dell EMC Navisphere™ or NaviCLI.

Note: Performance could be impacted on a VNX series or CLARiiON storage system if many hosts are using MPIO and a SP reboot is performed, which causes all the LUNs to failover to a particular SP.

To perform a trespass of all LUNs not on their default SP, run the NaviCLI command on both SPs of the VNX series or CLARiiON storage system:

```
naviseccli -h ip_of_SPA trespass mine
naviseccli -h ip_of_SPB trespass mine
```

The lsvg command

The AIX **`lsvg`** command, when used with the **`-p`** flag, displays devices in use by the specified volume group. This command, however, is not designed to operate with PowerPath or with storage--system logical devices that are addressable as different hdisk devices. In general, the output of **`lsvg -p vgname`** shows correct information, but several administrative tasks change the ODM and could cause `lsvg` to show misleading information. These tasks include:

- Use of the `pprootdev` tool. This tool changes the ODM and is intended to be used when you expect to reboot the system soon after using `pprootdev`. The **`lsvg`** command shows misleading device information when run after `pprootdev`. This is not an indication that something is wrong. A reboot corrects the `lsvg` output but it is not required.
- Use of `cfgmgr` to create new hdisk devices after PowerPath is already configured. Always run `powermt config` after adding new devices to include them in PowerPath's configuration.

Actions that use bosboot

After a system boots from a PowerPath device, the **`bosboot`** command cannot function correctly. This happens because of the state of the configuration after booting from a PowerPath device and the fact that `bosboot` expects the boot device to be an hdisk, not an hdiskpower device. Several system administrative tasks require the system boot image to be rebuilt, and these will fail if `bosboot` cannot run successfully. These tasks include applying certain software patches and using the `mksysb` utility to create a system backup. To address this limitation, the `pprootdev` tool provides a `fix` option that corrects the configuration to allow `bosboot` to work. Run **`pprootdev fix`** before undertaking any administrative task that runs **`bosboot`**. This corrects the configuration for `bosboot` but does not change the PowerPath boot switch; the next system boot still uses PowerPath. You need to run **`pprootdev fix`** only once after a system boots using PowerPath; after that, **`bosboot`** should function correctly until the system boots again.

Special recovery procedures: Path failure during bootup

For most situations involving error recovery, refer to the PowerPath manual to seek resolution. The only situation that can occur when booting over fibre (that may not be covered in the PowerPath manual) involves booting the system with a failed path. If the system boots up with a failed path, all devices that were available down the failed path will show up as Defined. In this situation, once the problem with the path has been corrected, the following procedure applies to VNX series, CLARiiON, and Symmetrix storage systems:

1. To determine which adapters have failed, use the commands in the following example. From this output we know that the adapter at 20-68 had the problem.

```
(1)(AIX4.3)> lsdev -Cc disk | grep Defined
hdisk4       Defined   20-68-01      EMC Symmetrix FCP Raid1
hdisk6       Defined   20-68-01      EMC Symmetrix FCP Raid1
hdisk7       Defined   20-68-01      EMC Symmetrix FCP Raid1
hdisk5       Defined   20-68-01      EMC Symmetrix FCP Raid1
hdisk8       Defined   20-68-01      EMC Symmetrix FCP Raid1
hdisk9       Defined   20-68-01      EMC Symmetrix FCP Raid1
hdisk10      Defined   20-68-01      EMC Symmetrix FCP Raid1
hdisk11      Defined   20-68-01      EMC Symmetrix FCP Raid1
hdisk12      Defined   20-68-01      EMC Symmetrix FCP Raid1
hdisk13      Defined   20-68-01      EMC Symmetrix FCP Raid1
hdisk14      Defined   20-68-01      EMC Symmetrix FCP Raid1
hdisk15      Defined   20-68-01      EMC Symmetrix FCP Raid1
hdisk16      Defined   20-68-01      EMC Symmetrix FCP Raid1

(0)(AIX4.3)> lsdev -Cc adapter | grep fcs
fcs0    Available 20-68    FC Adapter
fcs1    Available 40-58    FC Adapter
```

2. Once the problem with the failed path has been corrected, place the devices that are in a Defined state into an Available state using the **cfgmgr** command with the **–vl** parameters. Do not use the emc_cfgmgr script. As an alternative, the **mkdev –l** command can be used as well.

   ```
   (0)(AIX4.3)> cfgmgr -vl fcs0
   ```

3. Verify that those devices that were in a Defined state are now in the Available state using the **lsdev** command.

```
(0)(AIX4.3)l> lsdev -Cc disk | grep 20-68
hdisk4       Available 20-68-01      EMC Symmetrix FCP RAID1
hdisk6       Available 20-68-01      EMC Symmetrix FCP RAID1
hdisk7       Available 20-68-01      EMC Symmetrix FCP RAID1
hdisk5       Available 20-68-01      EMC Symmetrix FCP RAID1
hdisk8       Available 20-68-01      EMC Symmetrix FCP RAID1
hdisk9       Available 20-68-01      EMC Symmetrix FCP RAID1
hdisk10      Available 20-68-01      EMC Symmetrix FCP RAID1
hdisk11      Available 20-68-01      EMC Symmetrix FCP RAID1
hdisk12      Available 20-68-01      EMC Symmetrix FCP RAID1
hdisk13      Available 20-68-01      EMC Symmetrix FCP RAID1
hdisk14      Available 20-68-01      EMC Symmetrix FCP RAID1
hdisk15      Available 20-68-01      EMC Symmetrix FCP RAID1
hdisk16      Available 20-68-01      EMC Symmetrix FCP RAID1
```

4. If you use the **lspv** command at this time, you may notice that the PVIDs for the now-available hdisks may reside on the underlying hdisk devices instead of the hdiskpower devices. Run the **powermt config** command to place all of the PVIDs back on to the hdiskpower devices.

   ```
   (0)(AIX4.3)> powermt config
   ```

5. Run **powermt restore** to bring the devices back online if necessary and available for PowerPath for load balanced I/O.

   ```
   (0)(AIX4.3)> powermt restore
   ```

rootvg resides in Symmetrix storage array but mirrored across PowerPath adapters

In some situations it may be desired to not place rootvg under PowerPath control, even though you are booting from the storage array. To help protect your environment from path failure in this scenario rootvg can be mirrored across two different disks across your Powerpath controlled HBA(s).

Recommendations

- AIX operating system installed and booting properly from the Symmetrix storage array (one cable connected)
- An approved PowerPath configuration
- Most recent Dell EMC ODM definitions for the appropriate OS level of AIX
- PowerPath Version 3.0 or higher with a valid license

As with any procedure involving rootvg, an mksysb should be performed to back up any critical data. To make sure that you understand the configuration used, included is a brief explanation of the complete configuration used for this section of the documentation. This configuration has a total of eight hdisks. Four belong to the adapter at location code 10-58-01, and four belong to the adapter at 40-60-01. If you look at the output from the second command

(**inq**), which is short for inquiry, you will see that devices 01A and 019 appear only once, whereas devices 01B, 01C, and 01D appear twice. You can see that the set of devices that appear twice will be under the control of PowerPath, where the set that appears once will not be. The first set of devices will contain `rootvg` and instead of being controlled by PowerPath they will be protected by LVM mirroring.

```
(0)(AIX5.1)> lsdev -Cc disk
hdisk1 Available 10-58-01 EMC Symmetrix FCP Raid1
hdisk0 Available 40-60-01 EMC Symmetrix FCP Raid1
hdisk2 Available 10-58-01 EMC Symmetrix FCP Raid1
hdisk3 Available 10-58-01 EMC Symmetrix FCP Raid1
hdisk4 Available 10-58-01 EMC Symmetrix FCP Raid1
hdisk5 Available 40-60-01 EMC Symmetrix FCP Raid1
hdisk6 Available 40-60-01 EMC Symmetrix FCP Raid1
hdisk7 Available 40-60-01 EMC Symmetrix FCP Raid1

(127)(AIX5.1)>inq
Inquiry utility, Version V7.2-154 (Rev 16.0)      (SIL Version V4.2-154)
Copyright (C) by EMC Corporation, all rights reserved.
For help type inq -h.

........

---------------------------------------------------------------------
DEVICE       :VEND    :PROD            :REV   :SER NUM    :CAP(kb)
---------------------------------------------------------------------
/dev/rhdisk0 :EMC     :SYMMETRIX       :5568  :0901A140   :2208960
/dev/rhdisk1 :EMC     :SYMMETRIX       :5568  :09019130   :2208960
/dev/rhdisk2 :EMC     :SYMMETRIX       :5568  :0901B130   :2208960
/dev/rhdisk3 :EMC     :SYMMETRIX       :5568  :0901C130   :2208960
/dev/rhdisk4 :EMC     :SYMMETRIX       :5568  :0901D130   :2208960
/dev/rhdisk5 :EMC     :SYMMETRIX       :5568  :0901B140   :2208960
/dev/rhdisk6 :EMC     :SYMMETRIX       :5568  :0901C140   :2208960
/dev/rhdisk7 :EMC     :SYMMETRIX       :5568  :0901D140   :2208960
```

Recommendations

1. Use the commands shown in the following example to verify your configuration by making sure that you have two unique disks allocated to a minimum of two unique Fibre Channel paths. It is crucial that these devices are unique as they will not be under the control of the PowerPath product, and they must be unique to allow for LVM mirroring.

```
(0)(AIX5.1)>lsdev -Cc disk
hdisk1 Available 10-58-01 EMC Symmetrix FCP Raid1
hdisk0 Available 40-60-01 EMC Symmetrix FCP Raid1

(0)(AIX5.1)>inq
Inquiry utility, Version V7.2-154 (Rev 16.0)      (SIL Version V4.2-154)
Copyright (C) by EMC Corporation, all rights reserved.
For help type inq -h.

..

---------------------------------------------------------------------
DEVICE       :VEND    :PROD            :REV   :SER NUM    :CAP(kb)
---------------------------------------------------------------------
/dev/rhdisk0 :EMC     :SYMMETRIX       :5568  :0901A140   :2208960
/dev/rhdisk1 :EMC     :SYMMETRIX       :5568  :09019130   :2208960
```

Notice that there are currently two disks available, hdisk0 and hdisk1, each available down two different physical paths. The output from the **inq** command shows that in the field labeled SERIAL NUM there are two unique devices, one device 01A and the other device 019. Notice as well that the capacities are the same.

2. Since the installation started out by having a properly installed Fibre Channel boot system, we now have to extend `rootvg` to include the hdisk to which we will mirror. In this configuration, the installation was performed to hdisk1 so `rootvg` will be extended onto hdisk0.

```
(0)(AIX5.1)>extendvg rootvg hdisk0
0516-1254 extendvg: Changing the PVID in the ODM.
0516-014 linstallpv: The physical volume appears to belong to another
        volume group.
000c53cd00004c00
0516-631 extendvg: Warning, all data belonging to physical
        volume hdisk0 will be destroyed.
extendvg: Do you wish to continue? y(es) n(o)? yes
```

3. Tell the LVM to create a mapped mirror onto the new destination hdisk.

```
(0)(AIX5.1)>mirrorvg -m rootvg hdisk0
0516-1124 mirrorvg: Quorum requirement turned off, reboot system for this
        to take effect for rootvg.
0516-1126 mirrorvg: rootvg successfully mirrored, user should perform
        bosboot of system to initialize boot records. Then, user must modify
        bootlist to include:  hdisk0 hdisk1.
```

4. Use the **bosboot** command to reinitialize the boot sectors of the hdisks:

```
(0)(AIX5.1)>bosboot -ad /dev/ipldevice

bosboot: Boot image is 14300 512 byte blocks.
```

5. Set and then verify the bootlist order using the **bootlist** command. Be sure to include any additional boot devices that are specific to your environment.

```
(0)(AIX5.1)>bootlist -m normal hdisk0 hdisk1

(0)(AIX5.1)>bootlist -m normal -o
hdisk0
hdisk1
```

6. Verify the mirroring. Because there are two hdisks in the mirror, there will be twice the number of physical partitions in use.

```
(0)(AIX5.1)>lsvg -l rootvg
   rootvg:
       LV NAME           TYPE       LPs   PPs   PVs    LV STATE      MOUNT POINT
       hd5               boot       2     4     2      closed/syncd  N/A
       hd6               paging     16    32    2      open/syncd    N/A
       hd8               jfslog     1     2     2      open/syncd    N/A
       hd4               jfs        2     4     2      open/syncd    /
       hd2               jfs        116   232   2      open/syncd    /usr
       hd9var            jfs        3     6     2      open/syncd    /var
       hd3               jfs        6     12    2      open/syncd    /tmp
       hd1               jfs        1     2     2      open/syncd    /home
       hd10opt           jfs        5     10    2      open/syncd    /opt
       lg_dumplv         sysdump    256   256   1      open/syncd    N/A
```

7. Install PowerPath according to the *PowerPath Installation Guide*. Because `rootvg` is not under the control of PowerPath, do not use the section explaining how to install PowerPath with external boot.

Special recovery procedures: Path failure

If one or more paths fail in a mirrored configuration, you actually have two restorations to perform at all times. You must always keep in mind that once PowerPath has been restored to its original functioning condition, you must also make sure that `rootvg` is synchronized as well.

1. One indication that `rootvg` is no longer synchronized will be errors in the error log similar to the following. Also, you can check at anytime by using the **`lsvg`** command. Notice that under the column LV STATE, the logical volume is opened but us marked stale.

```
(0)(AIX5.1> errpt | more
   IDENTIFIER   TIMESTAMP  T C   RESOURCE_NAME  DESCRIPTION
   EAA3D429     0502112902 U S LVDD       PHYSICAL PARTITION MARKED STALE
   EAA3D429     0502112902 U S LVDD       PHYSICAL PARTITION MARKED STALE
   EAA3D429     0502112902 U S LVDD       PHYSICAL PARTITION MARKED STALE
   EAA3D429     0502112902 U S LVDD       PHYSICAL PARTITION MARKED STALE
   EAA3D429     0502112902 U S LVDD       PHYSICAL PARTITION MARKED STALE
   EAA3D429     0502112902 U S LVDD       PHYSICAL PARTITION MARKED STALE

(0)(AIX4.3)l82bc102/> lsvg -l rootvg
   rootvg:
   LV NAME             TYPE       LPs      PPs   PVs     LV STATE      MOUNT POINT
   hd5                 boot       1     2     2    closed/syncd  N/A
   hd6                 paging     64    128   2    open/syncd    N/A
   hd8                 jfslog     1     2     2    open/stale    N/A
   hd4                 jfs        1     2     2    open/stale     /
   hd2                 jfs        67    134   2    open/stale     /usr
   hd9var              jfs        1     2     2    open/stale     /var
   hd3                 jfs        2     4     2    open/stale     /tmp
   hd1                 jfs        1     2     2    open/syncd     /home
```

2. Once the problem with the failed path has been fixed to correct this condition, the `rootvg` volume group can be synchronized using the **`varyonvg`** command:

   ```
   (0)(AIX5.1)>varyonvg rootvg
   ```

3. Always make sure that after a period of time, all of your root volume group partitions eventually return to `open/syncd` status.

Using MKSYSB for backup and restore

Prerequisites

For AIX 4.3, minimum level EMC ODM definitions kit is 4.3.3.4. For AIX 5.1 and 5.2, minimum ODM definition is 5.1.0.0.

1. Before backing up your system you are required to run **`pprootdev fix`**. If you do not execute this command, your backup process will fail right away.
2. After running **`pprootdev fix`**, back up your host using **`mksysb`**; use any command syntax that you would ordinarily use for this process.
3. Because the AIX OS is not currently designed to deal with duplicate paths to devices without third-party software support, before starting the restore process you must reduce the number of paths to your boot device to one.
4. Boot from your **`mksysb`** tape.

5. If your configuration has not changed, it is best to restore using defaults. If you do need to select a new for restoration, select it, and then proceed with the restore. As part of the restoration process, the host will reboot automatically.

   Note: During the first automated reboot that occurs after the `mksysb restore` is complete, several scripts are run to rebuild devices since by default `recover_devices = yes` is set. This includes the powerpath0 driver that is necessary for PowerPath to properly configure devices and also boot externally.

6. Log in to the host. Since during this reboot the powerpath0 driver necessary for operation was just created, it is necessary to reboot the host again. This will replace the exclusive disk reservation on `rootvg` with a group reserve, allowing multiple paths to access this device.
7. Add the remaining cables to the host, and then run **`cfgmgr -vl`** against each individual adapter as it is added.
8. Run **`powermt config`** to reintegrate all of your additional devices and paths back into PowerPath.
9. Run **`pprootdev fix`**.
10. Run **`bosboot -ad /dev/ipldevice`**.
11. Update the bootlist to include all of the hdisks containing `rootvg`.

SRDF

This section describes requirements and procedures for configuring and booting a pair of AIX hosts from a set of Symmetrix SRDF disks.

IMPORTANT

SRDF, when used in conjunction with Fibre Channel boot, is currently not supported with AIX 4.3.3 and AIX 5.1, and generally supported with AIX 5.2 and higher.

This installation method assumes that all minimum host and HBA firmware revisions, operating-system revisions, and Symmetrix code revisions documented have been followed. This installation method also assumes that you have an existing AIX installation on internal disks at the minimum levels, and that this copy of AIX will be cloned using `alt_disk_install` to the Symmetrix boot device. The administrator can then boot off of either the internal or external disks as needed for maintenance or troubleshooting. It is possible to perform these tasks without using internal disks, but not practical. Several problems can occur when performing upgrades or troubleshooting without them, in a disaster recovery situation.

For SRDF to function as a disaster recovery solution for a bootable rootvg image, the two physical systems that are used (local host and remote host) need to be similar in model (that is, chrp and mp), adapter type, and layout to ensure that the servers can both boot off the same rootvg volume group image.

Figure 2 on page 53 is an example of a basic SRDF configuration that will be used as a reference throughout this section.

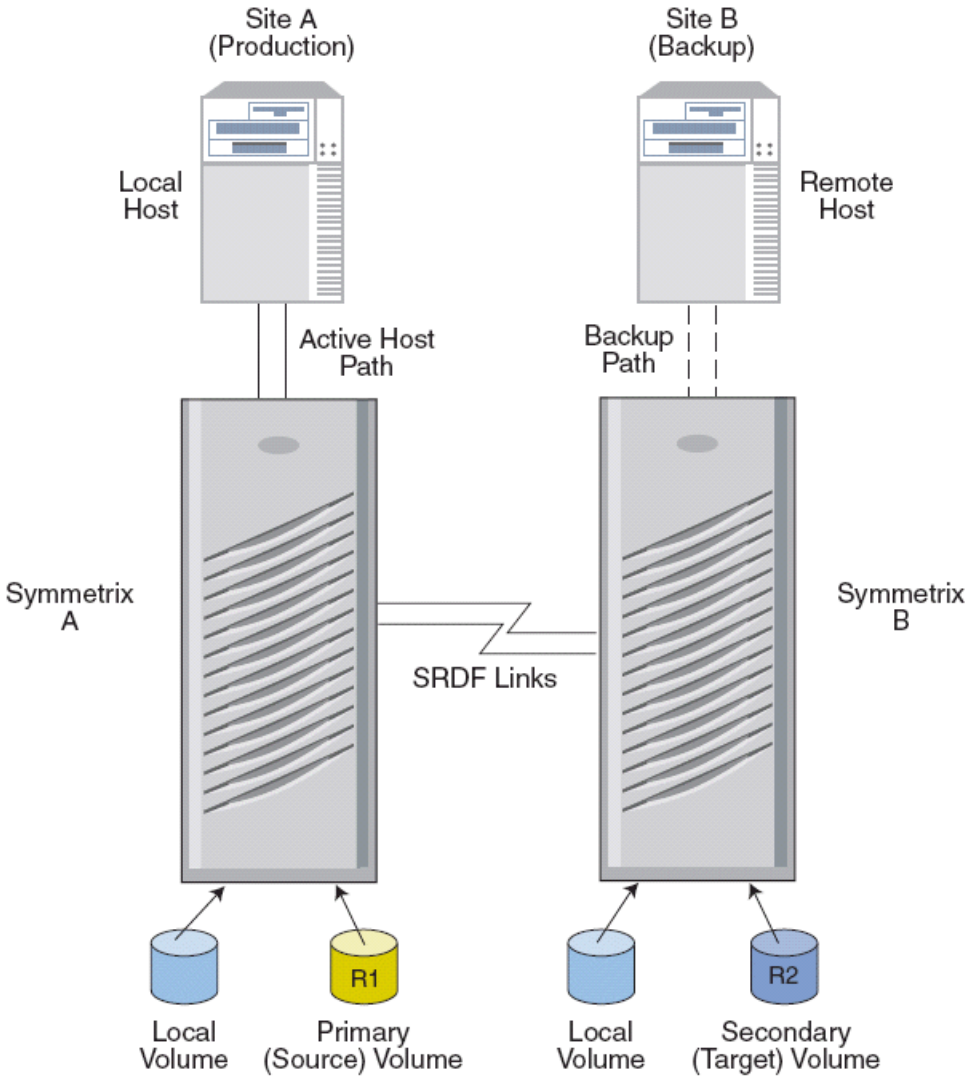


Figure 2 Basic SRDF configuration

SRDF installation overview

1. Identify the host(s) to be used for Fibre Channel boot.
2. Identify the type of host adapter to be used for Fibre Channel boot and install it.
3. Update the adapter microcode.
4. Update the system/service processor microcode.
5. Install the EMC ODM support package.
6. Attach the external disk
7. Install Solutions Enabler SYMCLI any license keys required.
8. Configure SRDF device groups.

9. Perform the clone from the current boot volume to the external disk using **`alt_disk_install`** (recommended) or perform a new installation.
10. Test cloned `rootvg` on local and remote hosts.

Identify the host(s) to be used for Fibre Channel boot

For SRDF to function as a disaster recovery solution for a bootable `rootvg` image, the two physical systems that are used (local host and remote host) need to be similar in model (i.e., chrp and mp), adapter type, and layout to ensure that the servers can both boot off the same rootvg volume group image.

Identify the type of host adapter to be used for Fibre Channel boot and install it

If the hosts do not have the adapter(s) installed, install only the adapter to be used for booting at this time. If the host already has multiple adapters installed, to avoid confusion you may want to remove or remove the cables from all fibre adapters except the one intended to be used as the boot adapter(s).

Install the EMC ODM support package

For AIX 5.2, complete the following steps:

1. From the FTP server, `ftp://ftp.emc.com/pub/elab/aix/ODM_DEFINITIONS`, download the most recent ODM fileset for the AIX version you are using. Currently, the fileset for AIX 5.2 is `EMC.AIX.5.2.0.X.tar.Z`. If a more recent fileset is available, use it as a substitute.
2. In the `/tmp` directory, uncompress the fileset. Substitute the appropriate filename revision for the command to work properly:

   ```
   uncompress EMC.AIX.5.2.0.X.tar.Z
   ```

3. After you uncompress the fileset, you will be left with a file with the `.tar` extension. Untar the resulting file. Substitute the appropriate filename revision for the command to work properly:

   ```
   tar -xvf EMC.AIX.5.2.0.X.tar
   ```

4. If the SMIT menu interface is preferred, invoke `smit installp` from the `/tmp` directory. Select **Install and Update from LATEST Available Software**.
5. Use the List function to select **`/tmp`** as the installation directory.
6. Use the **List** function to select **EMC Symmetrix AIX Support Software and EMC Symmetrix Fibre Channel Support Software.**
7. Press **Enter** *after* making all desired changes. You can use the default options.
8. Be sure to scroll all the way down to the bottom of the window to see the Installation Summary, and verify that the `SUCCESS` message is displayed.
9. Reboot the host for all changes to take effect.

Attach the external disk

1. Attach the external disk and the adapter into the SAN switch.
2. Verify that the adapters have logged in to the switch and the World Wide Names match the new adapters. These may be labeled on the back of the physical adapter, or can be found with the **`lscfg –pv | pg`** command and search for the `fcs#` stanza of the adapter you are using. The Network Address field in the `lscfg` output is equal to the World Wide Name.
3. Once you have attached the disks, run `cfgmgr` on the AIX host to add the newly discovered disks into the servers ODM. It is a good idea to only attach one physical connection at this time, even if you are later going to use a multipathing solution (e.g., PowerPath) due to problems that could occur during installation.
4. Zone the switch or disk array such that the machine being installed can only discover the disk(s) to be installed to. After the installation has completed, you can reopen the zoning so the machine can discover all necessary devices. Zone the R1 devices to the local host; zone the R2 devices to the remote host.
5. Assign PVIDs (physical volume identifiers) to all disks from the internal image of the AIX system that can access the disks. To do this, use the command **`chdev –l hdiskX –a pv=yes`**, where *X* is the appropriate number.
6. Create a table mapping PVIDs to physical disks.
7. Add the World Wide Name to the table with the following command:

   `lsattr -El hdisk# | grep ww_name`

8. Add LUN IDs to the table with the following commands:

   `lsattr -El hdisk# | grep lun_id`

   This information will be very helpful in determining the exact path and disk that you will be installing to from the boot/install menus or troubleshooting any problems. The PVIDs will be visible from the install menus by selecting the alternate attributes option when selecting the disk to install to. The World Wide Name and LUN IDs will be visible on the SMS multiboot bootlist. Do this from both servers because some values will be unique per server.

Example:

```
chdev -l hdisk2 -a pv=yes
chdev -l hdisk3 -a pv=yes
.
.
.

Local host disk details:

      Hdisk Number                    hdisk2
         Current Status =                Available
      Location =                      20-58-01
      Description =                   EMC Symmetrix FCP RDF1 Raid1
       World Wide Name =               0x5006048accc82de1
       Logical Storage array Number (LUN) ID = 0x0
      Volume Group Assigned = None
      Physical Volume (PVID) ID = 0002175f3043527f

      Hdisk Number =       hdisk3
```

```
      Current Status =        Available
      Location =              20-58-01
      Description =                 EMC Symmetrix FCP RDF1 Raid1
      World Wide Name =       0x5006048accc82de1
      Logical Storage array Number (LUN) ID = 0x1000000000000
      Volume Group Assigned = None
      Physical Volume (PVID) ID = 0002175f30435da6

Remote host disk details:

      Hdisk Number                  hdisk2
         Current Status =              Available
      Location =                    20-58-01
      Description =                 EMC Symmetrix FCP RDF2 Raid1
         World Wide Name =             0x5006048accc82dff
      Logical Storage array Number (LUN) ID = 0x10000000000000
      Volume Group Assigned =       None
      Physical Volume (PVID) ID =   0002175f3043527f

      Hdisk Number                  hdisk3
      Current Status =              Available
         Location =                    20-58-01
      Description =                 EMC Symmetrix FCP RDF2 Raid1
      World Wide Name =             0x5006048accc82dff
      Logical Storage array Number (LUN) ID = 0x11000000000000
      Volume Group Assigned =       None
      Physical Volume (PVID) ID =   0002175f30435da6
```

Verify that the external disks that you have selected to hold the external rootvg image are recognized to be bootable by the both the local and remote AIX servers. To do this, run the following command:

```
bootinfo -B hdisk#
```

The command will return a 1 if the disk is capable of booting that particular server. A return code of 0 means that you will not be able to boot off that disk device.

Install Solutions Enabler SYMCLI software

1. Install the Solutions Enabler SYMCLI software on both servers.
2. Add the `.profile` of the `root` user on local and remote AIX servers:

   ```
   export PATH=$PATH;/usr/symcli/bin
   export PATH=$PATH:/usr/lpp/EMC/Symmetrix/bin
   ```

3. Install license keys by running the `/usr/symcli/bin/symlmf` script and inserting the appropriate license keys.
4. Verify that the license keys are valid by running local commands on both local and remote hosts:
   - **`emc_cfgmgr`** – Detects new Symmetrix devices on the AIX servers.
   - **`symcfg discover`** – Builds or rebuilds the database file that stores device details.
   - **`symcfg list`** – Lists the symmetrix storage arrays that are visible to that server.
   - **`symcfg list -RA all`** – Shows the RDF Directors available to that server.
   - **`symrdf list`** – Shows state of all SRDF devices.

Configure SRDF device groups

An SRDF configuration has at least one source storage array and one target storage array. SRDF configurations may transfer data in a unidirectional or bidirectional manner. In a unidirectional configuration, all R1 devices reside in the source Symmetrix storage array and all R2 devices in the target Symmetrix storage array. Under normal operating conditions, data flows from the R1 devices to the target R2 devices.

The SRDF volumes should have been associated when your Dell EMC CE/SE configured the Symmetrix storage arrays. Device groups are system-specific, so it is a good idea to create group names that reflect the type activity you would perform on that group.

Create device groups for your boot devices, which will allow you to perform operations to all the devices at the same time. You may have one device group to hold the bootable `rootvg` image, and another device group to hold a database or mission-critical data. By using separate device groups, you can perform actions on them individually if necessary.

1. Create a device group to hold your disks containing the `rootvg` image:
   - on local server:

     ```
     symdg create prodrootvg –type RDF1
     ```

   - on remote server:

     ```
     symdg create prodrootvg -type RDF2
     ```

2. Add the devices you have selected for that external copy of rootvg. This example uses hdisk2 and hdisk3. Choose ones that are deemed bootable with the **bosboot –B** command.

```
symdev list -r1

Symmetrix ID: 000187900087

        Device Name   Directors                  Device
------------------- ------------ -------------------------------------
                                                                   Cap
Sym  Physical       SA :P DA :IT  Config        Attribute    Sts   (MB)
------------------- ------------ -------------------------------------

0000 /dev/rhdisk2   02C:1 01A:C0  RDF1+Mir      N/Grp'd      RW    4155
0001 /dev/rhdisk3   02C:1 16B:C0  RDF1+Mir      N/Grp'd      RW    4155
0002 /dev/rhdisk4   02C:1 16A:C1  RDF1+Mir      N/Grp'd      RW    4155
0003 /dev/rhdisk5   02C:1 01B:C1  RDF1+Mir      N/Grp'd      RW    4155


   Hdisk Number =    hdisk2
   Current Status =     Available
   Location =           20-58-01
   Description =                  EMC Symmetrix FCP RDF1 Raid1
      World Wide Name = 0x5006048accc82de1
   Logical Storage array Number (LUN) ID = 0x0
   Volume Group Assigned = none
   Physical Volume (PVID) ID = 0002175f3043527f

   Hdisk Number =    hdisk3
   Current Status =     Available
   Location =           20-58-01
   Description =                  EMC Symmetrix FCP RDF1 Raid1
      World Wide Name =  0x5006048accc82de1
   Logical Storage array Number (LUN) ID = 0x1000000000000
   Volume Group Assigned = none
   Physical Volume (PVID) ID = 0002175f30435da6
```

3. Add the disks to the `prodrootvg` device group with the following commands (perform on both local and remote hosts):
   - on local server (your RANGE value may be different):

     `symld -g prodrootvg –RANGE 000:001 addall dev`

   - on remote server (your RANGE value may be different):

     `symld -g prodrootvg -RANGE 010:011 addall dev`

4. You can now verify that the disks are now part of the `prodrootvg` group.

```
symdev list -r1

Symmetrix ID: 000187900087

        Device Name   Directors                  Device
------------------- ------------ ------------------------------------
                                                                  Cap
Sym  Physical       SA :P DA :IT  Config        Attribute    Sts   (MB)
------------------- ------------ ------------------------------------

0000 /dev/rhdisk2   02C:1 01A:C0  RDF1+Mir      Grp'd        RW    4155
0001 /dev/rhdisk3   02C:1 16B:C0  RDF1+Mir      Grp'd        RW    4155
0002 /dev/rhdisk4   02C:1 16A:C1  RDF1+Mir      N/Grp'd      RW    4155
0003 /dev/rhdisk5   02C:1 01B:C1  RDF1+Mir      N/Grp'd      RW    4155
```

You can also view specifics about the RDF devices in the group with the following command:

`symrdf -g prodrootvg query`

```
Device Group (DG) Name             : prodrootvg
DG's Type                          : RDF1
DG's Symmetrix ID                  : 000187900087

       Source (R1) View                Target (R2) View     MODES
--------------------------------    ------------------------ ----- ------------
             ST                  LI      ST
Standard      A                   N       A
Logical       T  R1 Inv   R2 Inv  K       T  R1 Inv   R2 Inv       RDF Pair
Device  Dev   E  Tracks   Tracks  S Dev   E  Tracks   Tracks MDA   STATE
-------------------------------- -- ------------------------ ----- ------------

DEV001  0000 RW       0        0 RW 0010 WD       0        0 S..  Synchronized
DEV002  0001 RW       0        0 RW 0011 WD       0        0 S..  Synchronized

Total          -------- --------           -------- --------
  Track(s)            0        0                  0        0
  MB(s)             0.0      0.0                0.0      0.0
```

Note: This output shows you the R1-to-R2 relationship as well as the State of the pair.

5. We can now establish the SRDF mirrors (if not performed previously) that will synchronize and overwrite the R2 side with an exact copy of the data of R1:

   `symrdf -g prodrootvg establish -full`

6. Depending if there is actual data on the R1 devices, this activity may take some time. To verify when the process is complete, perform the following:

symrdf -g prodrootvg -synchronized verify

If you receive the following message, the establish has completed:

```
All devices in the RDF group 'prodrootvg' are in the 'Synchronized'
state.
```

Perform the clone to the external disk using alt_disk_install (Recommended)

You are now ready to perform the `alt_disk_install` procedure to create a clone of the `rootvg` file system currently on the internal disks.

This example uses hdisk2 and hdisk3. Be careful not to select the Volume Logix Database Device that is represented as an hdisk on the AIX servers. In our example, it is shown as:

```
hdisk18 Available 20-58-01     EMC Symmetrix FCP RDF1 Raid1
```

You can verify which hdisk representation is your Volume Logix database using the following commands:

inq -dev /dev/rhdisk# -page0

```
-------------------------------------------------------------------------
BYTE 109:  VCMState  VCMDevice  GKDevice  MetaDevice  Shared
SA State      1         1          0         0          1
-------------------------------------------------------------------------
```

Look for the VCMDevice flag. If it is set to 1, then that is your Volume Logix device. Do not select that device by mistake when installing or booting.

If it is not already installed on your system, install the `bos.alt_disk_install` LPP from AIX CDs. Also, install the latest patch from IBM:

Installp -a -d/dev/cd0 bos.alt_disk_install

As an added safety precaution, make a backup `mksysb` of your progress up to this point.

smitty mksysb

Create the clone image using the **`alt_disk_install`** command:

```
Alt_disk_install -C hdisk2 hdisk3

Calling mkszfile to create new /image.data file.
Checking disk sizes.
Creating cloned rootvg volume group and associated logical volumes.
Creating logical volume alt_hd5
Creating logical volume alt_hd6
Creating logical volume alt_hd8
Creating logical volume alt_hd4
Creating logical volume alt_hd2
Creating logical volume alt_hd9var
Creating logical volume alt_hd3
Creating logical volume alt_hd1
Creating logical volume alt_hd10opt
Creating logical volume alt_dumplv
Creating /alt_inst/ file system.
Creating /alt_inst/home file system.
Creating /alt_inst/opt file system.
Creating /alt_inst/tmp file system.
Creating /alt_inst/usr file system.
Creating /alt_inst/var file system.
Generating a list of files
for backup and restore into the alternate file system...
Backing-up the rootvg files and restoring them to the
```

```
alternate file system...
Modifying ODM on cloned disk.
Building boot image on cloned disk.
forced unmount of /alt_inst/var
forced unmount of /alt_inst/usr
forced unmount of /alt_inst/tmp
forced unmount of /alt_inst/opt
forced unmount of /alt_inst/home
forced unmount of /alt_inst
forced unmount of /alt_inst
Changing logical volume names in volume group descriptor area.
Fixing LV control blocks...
Fixing file system superblocks...
Bootlist is set to the boot disk: hdisk2
```

Note: The bootlist has now been updated to hdisk2.

New installation to R1 disk (Optional)

If you choose not to use the alt_disk_install procedure you can perform a new installation from the AIX installation media (CD, tape or NIM image) using the below procedure:

1. Make sure that you have gathered the disk information for the specific disks you plant to install to. The installation process will display disk details similar to the following:

```
SCSI 4356 MB FC Harddisk id =@5006048accc82dff,10000000000000 ( slot=1 )
SCSI 4356 MB FC Harddisk id =@5006048accc82dff,11000000000000 ( slot=1 )
```

2. Put the installation media into the host and shut down. After the shutdown, you can remove the internal disk if desired (not recommended).
3. Boot from the installation media. If the installation media device was not part of the bootlist, you can select it by entering the SMS menus and selecting you installation device as you boot device.
4. Select the display to be used as the console during install.
5. Select the language to be used.
6. Select option 2, **Change/Show Installation Settings and Install**.
7. Select option 1, **System Settings**, and ensure that **New and Complete Overwrite** is selected.
8. If you have more than one disk allocated to the adapter, verify your installation disk by selecting option 77, **Display More Disk Information**, to see the PVID of the hdisks. Depending on the method that you used to identify the installation disk, as you continue selecting 77, additional information such as WWPN, SCSI ID, and LUN ID will appear. Also, on the initial disk selection screen, verify that **Yes** appears under the field marked bootable for the hdisk to which you will install. Once you have identified the correct hdisk device(s), select it and then continue. After the installation is complete, the host will automatically reboot from your external boot device.
9. Because you installed a new image, you will need to install the Dell EMC ODM support package to this new image (follow the procedure in "Installing the Dell EMC ODM support package" on page 40).

Test SRDF rootvg boot on local and remote hosts

SRDF control operation relationships

In a SRDF `rootvg` environment, there are two primary control operations to consider.

- **Split / establish** — A split operation stops remote mirroring between the R1 and R2 devices and makes R2 available to the remote host and R1 available to the local host. The establish operation would then **overwrite R2** with an exact copy of R1 and break the remote connection to R2 preventing any test data from interfering with production data.
- **Failover / failback** — A failover operation switches data processing from the R1 side to the R2 side. This disables the local host to R1 and the mirroring between the R1 and R2 side. It makes R2 available to the remote host. The failback operation will **Update the R1 copy** with changes made on the R2 side. This would be used in a true disaster recovery situation where you would not want to lose the changes made to the `rootvg` data.

You will now notice (if you performed an `alt_disk_install`) that you have an `altinst_rootvg` volume group on the external disks. The configuration is set to boot off of those disks during the next reboot of the local host.

```
Example:
lspv

hdisk0          00601910967f88f3                    rootvg
hdisk1          00015458bfe2e31f                    rootvg
hdisk2          0002175f3043527f                    altinst_rootvg
hdisk3          0002175f30435da6                    altinst_rootvg
```

Reboot the local host to activate the production `rootvg` on the external disk and verify that it can indeed boot off the cloned `rootvg` on the external devices.

```
Shutdown -Fr
```

After booting from the cloned installation, you will notice that the original installation is called `old_rootvg`.

```
lspv

hdisk0          00601910967f88f3                    old_rootvg
hdisk1          00015458bfe2e31f                    old_rootvg
hdisk2       0002175f3043527f             rootvg
hdisk3       0002175f30435da6             rootvg
```

All that must be performed now to boot from either copy is to change the bootlist to point to whichever copy of `rootvg` we may need. `Bootlist -m normal -o` lists the current boot device. `Bootlist -m normal hdisk#` changes the copy of `rootvg` from which you want to boot. If you have forgotten the original boot device, use the **`alt_disk_install -q`** command against any disk in an inactive copy of `rootvg` to find which device contains the bootable image.

```
alt_disk_install -q hdisk1
hdisk0  # (This is output from the above command.)
```

At this point, you have three copies of `rootvg` from which to boot: One copy on each server's internal disk, and a "floating" production copy on external disk. Next, boot off the local hosts internal disks again so we can move the cloned copy to the remote host and boot off of it:

```
bootlist -m normal -o
hdisk2   # (This is output from the above command.)
```

```
bootlist -m normal hdisk0

bootlist -m normal -o
hdisk0 # (This is output from the above command.)

shutdown -Fr
```

With both servers booted off their internal copies of `rootvg`, run the following command to fail over the R1 devices to the remote host. This is done to ensure that the shared `rootvg` copy contains the proper configuration in their ODM files for all devices. This will write-disable the local hosts connection to the R1, suspend RDF link traffic, and read/write-enable R2 to the remote host.

```
symrdf –g prodrootvg failover

Execute an RDF 'Failover' operation for device
group 'prodrootvg' (y/[n]) ? y

An RDF 'Failover' operation execution is
in progress for device group 'prodrootvg'. Please wait...

    Write Disable device(s) on SA at source (R1).............Done.
    Suspend RDF link(s).....................................Done.
    Read/Write Enable device(s) on RA at target (R2).........Done.

The RDF 'Failover' operation successfully executed for
device group 'prodrootvg'.
```

You should now see the state of the devices changed to “Failed over” in both the symrdf list and symrdf –g prodrootvg query commands.

symrdf list

```
Symmetrix ID: 000187900087
                              Local Device View
---------------------------------------------------------------------------
                  STATUS   MODES                       RDF  S T A T E S
Sym        RDF    --------- -----  R1 Inv   R2 Inv ----------------------
Dev  RDev  Typ:G  SA RA LNK  MDA   Tracks   Tracks Dev RDev Pair
---- ----  ------ --------- ----- -------  ------- --- ---- -------------

0000 0010   R1:2  WD RW NR   S..        0        0 WD  RW   Failed Over
0001 0011   R1:2  WD RW NR   S..        0        0 WD  RW   Failed Over


symrdf -g prodrootvg query

Device Group (DG) Name             : prodrootvg
DG's Type                          : RDF1
DG's Symmetrix ID                  : 000187900087

       Source (R1) View                Target (R2) View     MODES
--------------------------------    ------------------------ ----- ------------
            ST                  LI      ST
Standard     A                   N       A
Logical      T  R1 Inv   R2 Inv  K       T  R1 Inv   R2 Inv       RDF Pair
Device  Dev  E  Tracks   Tracks  S Dev   E  Tracks   Tracks MDA   STATE
-------------------------------- -- ------------------------ ----- ------------

DEV001  0000 WD       0        0 NR 0010 RW       0        0 S..   Failed Over
DEV002  0001 WD       0        0 NR 0011 RW       0        0 S..   Failed Over

Total          -------- --------           -------- --------
  Track(s)            0        0                  0        0
```

```
MB(s)              0.0      0.0                0.0      0.0
```

To avoid confusion, shut down the local host at this time to avoid a conflict on the network with two servers responding to the primary IP_LABEL of the local host. You can also use the `uname -a` output to keep track of the physical server you are on.

On the local host, issue the following commands:

```
>uname -a
AIX local 1 5 000154584C00
>shutdown -F
```

On the remote host, verify the proper boot device by using the PVID and LUN ID. Note that the World Wide Name will be different because of the adapter used on this server. This is the R2 device that mirrors the bootable R1 device on the local host.

```
Hdisk Number =        hdisk2
Current Status =      Available
Location =            20-58-01
Description =                  EMC Symmetrix FCP RDF2 Raid1
World Wide Name =     0x5006048accc82dff
Logical Storage array Number (LUN) ID = 0x10000000000000
Volume Group Assigned = None
Physical Volume (PVID) ID = 0002175f3043527f
```

In this example, the external bootable drive is defined as hdisk2 on both servers; however, the internal bootable disk on the remote host is hdisk1. Change the bootlist to boot of the external bootable disk and reboot the remote host.

```
bootlist -m normal -o
hdisk1   # (This is output from the above command.)

bootlist -m normal hdisk2

bootlist -m normal -o
hdisk2 # (This is output from the above command.)
uname -a
AIX remote 1 5 006019104C00

shutdown -Fr
```

Adding PowerPath

As with any procedure involving `rootvg`, a `mksysb` should be performed to back up any critical data. The installation procedure migrates volume groups built on Symmetrix devices from AIX hdisks to PowerPath hdiskpower devices. After installing PowerPath, you need to vary on the volume groups and mount any file systems that use those volume groups, but you do not need to reconfigure the volume groups themselves.

1. Note the adapters that will be used for the PowerPath configuration. Remember that only one of them should have any devices configured down it at this time. In this configuration, there are two adapters:

   ```
   > lsdev -Cc adapter | grep fcs

   fcs0         Available    20-58  FC Adapter
   fcs1         Available    20-60  FC Adapter
   ```

2. Verify that devices are configured down only one path at a time. You should see your hdisk devices with all of the location codes matching only one of the adapters you determined in step 1. The exception to this would be your internal hdisk if one were present. All of these hdisks, with the exception of the internal disk, are available for the adapter located at 20-58, which corresponds to fcs0 (this text has been truncated).

   ```
   > lsdev -Cc disk
   hdisk2  Available 20-58-01     EMC Symmetrix FCP RDF1 Raid1
   hdisk3  Available 20-58-01     EMC Symmetrix FCP RDF1 Raid1
   .
   .
   .
   ```

3. Install PowerPath as described in the *PowerPath for UNIX Installation Guide*. Refer to the section on AIX. It will guide you through properly installing the PowerPath software and checking your registration.

4. Because `rootvg` is varied on, there is an exclusive disk reserve placed on the device preventing it from being accessed down the second path. The next three steps are necessary to replace that exclusive disk reserve with a group reserve. The group reserve allows the device to be accessed down multiple paths. Start by using the PowerPath configuration command if you have not already done so.

   ```
   > powermt config
   powerpath0 created
   ```

5. Use the `pprootdev` tool to enable PowerPath protection of `rootvg`. The `pprootdev` tool is used to change AIX configuration rules and update the boot image so that the AIX Logical Volume Manager will use hiskpower devices to vary on the root volume group the next time the system is booted.

   ```
   > pprootdev on
   bosboot: Boot image is 8669 512 byte blocks.
   PowerPath boot is enabled for the next system boot
   ```

6. To actually place the group reserve on this device the host must now be rebooted.

   ```
   > shutdown -Fr
   ```

7. Once the system is back up, the `rootvg` device should have the group reservation set. You can now attach and configure the second path to the Symmetrix storage array at this time. Using the `emc_cfgmgr` script will configure automatically all paths and then update the PowerPath configuration.

   > **/usr/lpp/EMC/Symmetrix/bin/emc_cfgmgr**

8. Verify that all of your hdisks have been configured properly down each path. This process is covered in greater detail in the *PowerPath for UNIX Installation and Administration Guide*. (In this example, hdisk2 and hdisk23 are the same disk through different paths, represented by hdiskpower0).

```
> lsdev -Cc disk
…
hdisk2       Available 20-58-01     EMC Symmetrix FCP RDF1 Raid1
hdisk3       Available 20-58-01     EMC Symmetrix FCP RDF1 Raid1
…
hdiskpower0  Available 20-58-01     PowerPath Device
hdiskpower1  Available 20-58-01     PowerPath Device
…
hdisk23      Available 20-60-01     EMC Symmetrix FCP RDF1 Raid1
hdisk24      Available 20-60-01     EMC Symmetrix FCP RDF1 Raid1
…
```

9. Run the **pprootdev fix** command:

   > **pprootdev fix**

   You may now run **bosboot**.

   PowerPath boot remains enabled for the next system boot.

10. Run **bosboot** to reinitialize the boot record on the boot device:

    > **bosboot -ad /dev/ipldevice**

11. Update the bootlist to include all of the underlying hdisk devices of `rootvg`. Be sure to include any additional devices that you may need.

    > **bootlist -m normal hdisk2 hdisk23**

12. Reboot the host again:

    > **shutdown -Fr**

13. When the host is back online you can verify that you have booted from the PowerPath-protected `rootvg` by using the **lspv** command. Notice that `rootvg` now belongs to an hdiskpower device (this text has been truncated).

```
> lspv
…
hdiskpower0     00015458eb951002                    rootvg
hdiskpower1     00015458eb9524df                    rootvg
```

14. Use the **powermt display** command to check the status of the various paths to your disks:

```
>powermt display
```

```
Symmetrix logical device count=20
CLARiiON logical device count=0
================================================================================
----- Host Bus Adapters ---------  ------ I/O Paths -----  ------ Stats ------
### HW Path                        Summary   Total   Dead  IO/Sec Q-IOs Errors
================================================================================
  0 fscsi0                         optimal      20      0       -     0      0
  1 fscsi1                         optimal      16      0       -     0      0
```

Potential problems

This section shows examples of user interaction within an RS/6000 system. Be aware that a non-RS/6000 system may display similar, but somewhat different, results.

BOOT device LEDs

If the server won't boot or hangs at E1F5 or 20EE000B appears on the LED panel or other errors occur, the server can't find the boot device. The issue is related to the way the server uniquely identifies the drives. When you change anything in the path to the drive, the firmware recognizes this as a different hdisk. On the shared copy of `rootvg`, there is a definition for hdisk2, which is the R1 path to the drive. When you attempt to boot that image, it may conflict with the definition of the original R1 path that no longer exists. This is where the documentation is invaluable.

The external copy of `rootvg` is configured to boot off of hdisk2 through the World Wide Name path of `0x5006048accc82de1 LUN ID 0 (0x0)` in its normal state.

```
Hdisk Number =                  hdisk2
Current Status =                Available
Location =                      20-58-01
Description =                   EMC Symmetrix FCP RDF1 Raid1
World Wide Name =               0x5006048accc82de1
Logical Storage array Number (LUN) ID = 0x0
Volume Group Assigned =         none
Physical Volume (PVID) ID =     0002175f3043527f
```

When you split or fail over the shared copy of rootvg and boot the remote host, the path to the device will have changed. The AIX server firmware may decide that the change in the physical path is a different hdisk. In our example, it is:

```
SCSI 4356 MB FC Harddisk id=@5006048accc82dff,10000000000000 ( slot=1 )
```

This matches our remote host hdisk2 definition from earlier.

```
Hdisk Number                    hdisk2
Current Status =                Available
Location =                      20-58-01
Description =                   EMC Symmetrix FCP RDF2 Raid1
World Wide Name =               0x5006048accc82dff
Logical Storage array Number (LUN) ID =   0x10000000000000
Volume Group Assigned =         None
Physical Volume (PVID) ID =     0002175f3043527f
```

To resolve, boot to the SMS screen by pressing a **F1** on a GUI (or type **`1`** on a tty) on the console before the fifth keyword appears on the banner screen:

```
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
```

```
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000


memory      keyboard     network      scsi



<type ìF1î on a graphics Keyboard or ì1î on a tty>


RS/6000 Firmware
Version WIL03115
(c) Copyright IBM Corp. 2000  All rights reserved.
-------------------------------------------------------------------------------
System Management Services

1  Display Configuration
2  Multiboot
3  Utilities
4  Select Language


 <select ì2î to enter the Multiboot menu>

RS/6000 Firmware
Version WIL03115
(c) Copyright IBM Corp. 2000  All rights reserved.


-------------------------------------------------------------------------------
Multiboot
1  Select Software
2  Software Default
3  Install From
4  Select Boot Devices
5  OK Prompt
6  Multiboot Startup <OFF>

<select ì4î to Select Boot Devices>
```

Select (or reselect) the proper WWN and LUN ID that you wrote down as the first boot device for that server.

```
1  SCSI 4356 MB FC Harddisk id=@5006048accc82dff,10000000000000 ( slot=1 )
```

Exit your way back through the menus with ìxî and continue the boot process.

```
===>X
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000                                 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000       STARTING SOFTWARE         RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000         PLEASE WAIT...          RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000                                 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
```

```
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
-------------------------------------------------------------------------------
                                Welcome to AIX.
                       boot image timestamp: 14:20 10/28
                 The current time and date: 17:00:56 10/28/2003
                number of processors: 2    size of memory: 256Mb
boot device: /pci@fee00000/fibre-channel@b/disk@5006048accc82dff,10000000000000:
2
closing stdin and stdout...
```

After a successful boot, you should now see the internal disks shown as `old_rootvg`. Note that our configuration renamed the PVID `0002175f3043527f` as `hdisk19` and our old definition of hdisk2 is in a Defined state. In this copy of the `rootvg`, hdisk2 is the path from the primary server to the R1 device. The WWN and possibly the LUN ID will be different, but the PVID should be the same.

```
Hdisk Number =                   hdisk19
Current Status =                 Available
Location =                       20-58-01
Description =                    EMC Symmetrix FCP RDF2 Raid1
World Wide Name =                0x5006048accc82dff
Logical Storage array Number (LUN) ID =   0x10000000000000
Volume Group Assigned =          rootvg
Physical Volume (PVID) ID =      0002175f3043527f

>lspv

hdisk0          00601910967f88f3                    old_rootvg
hdisk1          00015458bfe2e31f                    old_rootvg
hdisk19         0002175f3043527f                    rootvg
hdisk20         0002175f30435da6                    rootvg
```

You may also get errors running Symmetrix commands due to the path differences.

```
>symrdf -g prodrootvg query

Error opening the gatekeeper device for communication to the Symmetrix
```

If this happens, run **`symcfg discover`** to dynamically rediscover and repair the connections.

```
>symcfg discover
This operation may take up to a few minutes. Please be patient...

>symrdf -g prodrootvg query

Device Group (DG) Name             : prodrootvg
DG's Type                          : RDF1
DG's Symmetrix ID                  : 000187900087
Source (R1) View                 Target (R2) View     MODES
--------------------------------    ------------------------ ----- ------------
             ST                  LI      ST
Standard      A                   N       A
Logical       T  R1 Inv   R2 Inv  K       T  R1 Inv   R2 Inv       RDF Pair
Device  Dev   E  Tracks   Tracks  S Dev   E  Tracks   Tracks MDA   STATE
-------------------------------- -- ------------------------ ----- ------------
DEV001  0000 WD       0        0 NR 0010 RW      221        0 S..   Failed Over
DEV002  0001 WD       0        0 NR 0011 RW       63        0 S..   Failed Over

Total          -------- --------           -------- --------
```

```
  Track(s)            0        0                  284        0
  MB(s)             0.0      0.0                  8.9      0.0
```

Note: There are now invalid R1 tracks due to the changes during the boot cycle.

552/554 LED

LED 552 recovery procedure

There could be hangs during a boot at a 552 LED if you don't have the disks in an available state to that server (i.e., in failovermode, but trying to boot from the local server). Make sure that the SRDF disks are in the proper state for access from the server from which you are booting. Other 552 LED hangs can be resolved by booting the server into maintenance mode, and removing all the disk definitions in rootvg.

1. Insert CD of the same AIX version into the CD/DVD reader, and Boot into System Maintenance Mode and choose option F5 or 5.

   Boot to the SMS screen by typing an F5 on a GUI (or 5 on a tty) on the console before the 5th keyword appears on the banner screen:

```
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000


memory   keyboard  network  scsi -> <type "F5" on a graphics Keyboard or "5" on a tty>
when this options start to show on screen
```

2. Choose the following options:

   Access the root volume group.

   - Choose option # 2. choose this console
   - Choose option # 1. type 1 and press **Enter** to have English during install
   - Choose option # 1. start maintenance mode for system recovery
   - Choose option # 1. access a root volume group
   - Choose 0 to continue
   - Choose option # 1. access this volume group and start a shell

Note: To choose the boot logical volume to be mounted look for the size of the volume which should match the size of the standard and then check for ìhd5î after you pick the device – if the device shows anything different than hd5 hit 99 to go back and pick another device.

3. Remove hdisks defined for rootvg volume group:
   - Run **`lspv`** and record all hdisks in rootvg
   - Run **`rmdev -dl hdiskxx`** for all hdisks on rootvg – this step removes the hdisk definition on the system
   - Run **`lspv`** again to verify the hdisks were removed.
4. Then reboot the server:

   ```
   > shutdown -Fr
   ```

This procedure removes the references of the original hdisks from the ODM and forces the boot sequence to examine the disks during the boot sequence. Secondary host should boot up normally.

LED 554 SMS recovery procedure

If you receive a 554 LED, it may be caused by the multiple paths to your boot disk. As of this writing, AIX 5.1 and older versions do **not support** multiple paths to the boot device. If more than one path is configured to the boot device, remove all additional paths in order to keep only *one* path configured, or physically remove the cables from the back of the adapter prior to booting.

To resolve, after removing all but one path to the boot disk, boot to the SMS screen by pressing **F1** on a GUI (or type **1** on a tty) on the console before the fifth keyword appears on the banner screen:

```
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
```

```
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000

memory      keyboard     network      scsi



<type "F1" on a graphics Keyboard or "1" on a tty>


RS/6000 Firmware
Version WIL03115
(c) Copyright IBM Corp. 2000  All rights reserved.
-------------------------------------------------------------------------------
System Management Services

1  Display Configuration
2  Multiboot
3  Utilities
4  Select Language

 <select "2" to enter the Multiboot menu>

RS/6000 Firmware
Version WIL03115
(c) Copyright IBM Corp. 2000  All rights reserved.
-------------------------------------------------------------------------------
Multiboot

1  Select Software
2  Software Default
3  Install From
4  Select Boot Devices
5  OK Prompt
6  Multiboot Startup <OFF>


<select "4" to Select Boot Devices>
```

Select the proper WWN and LUN ID that you wrote down as the boot device for the remove (R2) copy of the boot image.

```
1  SCSI 4356 MB FC Harddisk id=@5006048accc82dff,0x10000000000000 ( slot=1 )
```

Exit your way back through the menus with "x" and continue the boot process.

```
===>X
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000                                 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000       STARTING SOFTWARE         RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000         PLEASE WAIT...          RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000                                 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
```

```
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000
RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000 RS/6000


-------------------------------------------------------------------------------
                                Welcome to AIX.
                       boot image timestamp: 14:20 10/28
                 The current time and date: 17:00:56 10/28/2003
                number of processors: 2    size of memory: 256Mb
boot device: /pci@fee00000/fibre-channel@b/disk@5006048accc82dff, 0x10000000000000:
2
closing stdin and stdout...
```

Upon completion of a successful boot, verify that all disks are available.

LED 554 bootable media recovery procedure

1. Boot in Maintenance Mode (from NIM in this example). You will be prompted to select Root Volume Group.

```
           Access a Root Volume Group

Type the number for a volume group to display the logical volume information
and press Enter.

   1)   Volume Group 0052904a00004c00000000fba68bb032 contains these disks:
          hdisk0  17357 10-60-00-0,0
   2)   Volume Group 0052904a00004c00000000fc5ac13c3a contains these disks:
          hdisk2  21502     20-58-01













    Choice: 2
```

2. Select the entry corresponding to your `rootvg`, (check VGID).

   Notice that **`hdisk2`** is the device configured by the AIX kernel (Maintenance Mode) in this example.

```
                        Volume Group Information

--------------------------------------------------------------------------------
   Volume Group ID 0052904a00004c00000000fc5ac13c3a includes the following
   logical volumes:

        hd5         hd6         hd8         hd4         hd2      hd9var
        hd3         hd1     hd10opt        lv00        lv01
--------------------------------------------------------------------------------




Type the number of your choice and press Enter.

   1) Access this Volume Group and start a shell
   2) Access this Volume Group and start a shell before mounting filesystems

  99) Previous Menu

    Choice [99]: 1
```

3. Select **Access this Volume Group and start a shell**.
4. Once `rootvg` is imported, `lspv` shows the following volumes

```
#lspv
hdisk0          0052904a0c2d3952    old_rootvg
hdisk1          none                none
hdisk3          0052904a5ac12d8a    rootvg
hdisk4          none                none
hdisk5          none                none
hdisk6          0052904a5ac12d8a    rootvg
hdisk7          none                none
hdisk8          none                none
```

Notice that hdisk2 is missing. This is because the **`lspv`** command gets the info from the ODM on the disk. According to the previous display, `rootvg` belongs to hdisk3 and hdisk6 on ODM.

5. Check the boot logical volume

```
# lslv -m hd5
hd5:N/A
LP     PP1  PV1               PP2  PV2               PP3  PV3
0001   0001 hdisk3
```

Notice that hd5 belongs to hdisk3.

6. Remove the additional path (hdisk6 in this example):

```
# rmdev -dl hdisk6
hdisk6 deleted
```

7. Execute the **bosboot** command, but be sure to execute it against the correct device.

Remember that the `rootvg` disk configured in Maintenance mode is hdisk2; the system has also configured a special device file `/dev/ipldevice` pointing to hdisk2.

Consequently, the **`bosboot`** command must be executed on `/dev/ipldevice`.

If you select hdisk3 (device associated with hd5), the **`bosboot`** command will fail with the following error message:

```
# bosboot -ad /dev/hdisk3
odmget: Could not retrieve object for CuAt, odm errno 5904

bosboot: Boot image is 18040 512 byte blocks.
#
```

Execute the **`bosboot`** command on `/dev/ipldevice`:

```
# bosboot -ad /dev/ipldevice
bosboot: Boot image is 18040 512 byte blocks.
```

8. Once the boot image is successfully created, reboot the host:

   ```
   # shutdown -Fr
   ```

Special considerations

This section contains special considerations to be aware of.

Special considerations when PowerPath protection of rootvg is enabled

As explained before, if you receive a 552/554 LED when trying to boot off the R2 copy of `rootvg`, it may be caused by the multiple paths to your boot disk.

It may also be caused to the fact that `pprootdev on` (enabled) and the PVID of rootvg disk is associated with hdiskpower device.

With AIX 5.1, the PVID of `rootvg` *must* be associated with hdisk even if PowerPath protection of `rootvg` is enabled (`pprootdev on`).

To resolve this issue, the **`pprootdev fix`** command must be integrated into a script called by `/etc/inittab` at boot time or run after every boot. This will ensure that PVID is always associated with hdisk(s).

Special considerations when data volume groups are configured on hdiskpower devices

When data volume groups are configured on R1 devices under PowerPath control, these volume groups may not be activated automatically in case of failover to the R2 side.

This is because the hdiskpower PVID entry of R1 device remains configured in the ODM when a failover is executed. Furthermore, when booting off the R2 side, new hdiskpower devices will be configured for data disks. That results in duplicate PVID entries, as shown next:

```
# odmget -q 'value like *0030379ba4960689* and attribute=pvid' CuAt
CuAt:
        name = "hdiskpower4"
        attribute = "pvid"
        value = "0030379ba4960689000000000000000"
        type = "R"
        generic = "D"
        rep = "s"
        nls_index = 2
 CuAt:
```

```
        name = "hdiskpower5"
        attribute = "pvid"
        value = "0030379ba4960689000000000000000"
        type = "R"
        generic = "D"
        rep = "s"
        nls_index = 2

# lspv
...
hdiskpower0     0052904a5ac12d8a                    rootvg
hdiskpower3     0030379b63cb7a4c                    None
hdisk5          none                                None
hdisk10         none                                None
hdiskpower5     0030379ba4960689                    datavg

# lsdev -Cc disk
...
hdiskpower4 Defined   3p-08-01      PowerPath Device
hdiskpower5 Available 4k-08-01      PowerPath Device
```

To fix this issue, remove the hdiskpower entry in Defined state:

```
# rmdev -dl hdiskpower4
hdiskpower4 deleted
```

Then, you can **varyonvg** the data volume group and mount the file systems:

```
# varyonvg datavg

# lsvg -o
datavg
rootvg
```

Failure recovery scenarios

This section contains failure recovery scenarios:

Scheduled shutdown of production local server

During normal procedures in this example we have the local server booting off of the production copy of `rootvg`. To intentionally move this over to the remote server for maintenance or scheduled outage, perform the following steps:

- On the local server:

  Stop any applications and perform a shutdown to stop all processes and close files.

  >**shutdown -F now**

- On remote server:

  If you are planning to retain the changes made on the rootvg during this outage, perform a "failover" operation. If you are going to discard the changes perform a "split" operation. This will stop the synchronized traffic from the R1 to the R2 side and make the R2 disk available to the remote host.

```
> symrdf -g prodrootvg failover
Execute an RDF 'Failover' operation for device group 'prodrootvg' (y/[n]) ? y

An RDF 'Failover' operation execution is in progress for device group 'prodrootvg'. Please
wait...

    Write Disable device(s) on SA at source (R1).............Done.
    Suspend RDF link(s).....................................Done.
    Read/Write Enable device(s) on RA at target (R2).........Done.

The RDF 'Failover' operation successfully executed for device group 'prodrootvg'.
```

Change the bootlist from the active rootvg on the remote server to the production R2 disk

>**bootlist -m normal hdisk2 hdisk3**

Reboot the remote server off of the production `rootvg`.

>**shutdown -Fr**

Power loss or hang of production local server

- On the local server:

  Record any LEDs or obtain any information that may help diagnose the cause of the problem.

- On the remote server:

  If you are planning to retain the changes made on the rootvg during this outage, perform a failover operation. If you are going to discard the changes, perform a split operation. This will stop the synchronized traffic from the R1 to the R2 side and make the R2 disk available to the remote host.

```
> symrdf -g prodrootvg failover
Execute an RDF 'Failover' operation for device group 'prodrootvg' (y/[n]) ? y
```

```
An RDF 'Failover' operation execution is in progress for device group 'prodrootvg'. Please
wait...

    Write Disable device(s) on SA at source (R1).............Done.
    Suspend RDF link(s).....................................Done.
    Read/Write Enable device(s) on RA at target (R2).........Done.

The RDF 'Failover' operation successfully executed for device group 'prodrootvg'.
```

Change the bootlist from the active rootvg on the remote server to the production R2 disk:

```
>bootlist -m normal hdisk2 hdisk3
```

Reboot the remote server off of the production rootvg:

```
>shutdown -Fr
```

Sysdump of production local server

If the local server performs a system dump, you may not be able to diagnose the system dump on the remote host. You may have to fail back to the original server to run analysis tools to debug the error.

- On the local server:

  Record any LEDs or obtain any information that may help diagnose the cause of the problem.

- On the remote server:

  If you are planning to retain the changes made on the rootvg during this outage, perform a failover operation. If you are going to discard the changes, perform a split operation. This will stop the synchronized traffic from the R1 to the R2 side and make the R2 disk available to the remote host.

```
> symrdf -g prodrootvg failover
Execute an RDF 'Failover' operation for device group 'prodrootvg' (y/[n]) ? y

An RDF 'Failover' operation execution is in progress for device group 'prodrootvg'. Please
wait...

    Write Disable device(s) on SA at source (R1).............Done.
    Suspend RDF link(s).....................................Done.
    Read/Write Enable device(s) on RA at target (R2).........Done.

The RDF 'Failover' operation successfully executed for device group 'prodrootvg'.
```

Change the bootlist from the active `rootvg` on the remote server to the production R2 disk:

```
>bootlist -m normal hdisk2 hdisk3
```

Reboot the remote server off of the production `rootvg`:

```
>shutdown -Fr
```

Extended outages

There may be times when the local server would be unavailable for an extended period of time. Consider using the swap function to reverse the roles of the R1 and R2 sides in this situation. That way, you can keep your local Symmetrix system synchronized during the production activity during the failover, and the delay of synchronizing the R1 copy with all the R2 updates is reduced.

This procedure assumes that the original R1 Symmetrix system is functioning properly and that you want to maintain the production copy of the data that is being updated on the R2 side in a failover situation.

Current state assumes a failover has been previously issued. You should see the state of the devices changed to `Failed over` in both the **`symrdf list`** and **`symrdf -g prodrootvg`** query commands.

`symrdf list`

```
Symmetrix ID: 000187900087
                                Local Device View
-------------------------------------------------------------------------
                  STATUS   MODES                         RDF  S T A T E S
Sym        RDF    --------- -----  R1 Inv   R2 Inv ----------------------
Dev  RDev  Typ:G  SA RA LNK  MDA   Tracks   Tracks Dev RDev Pair
---- ----  ------ --------- ----- -------  ------- --- ---- -------------

0000 0010   R1:2  WD RW NR   S..        0        0 WD  RW   Failed Over
0001 0011   R1:2  WD RW NR   S..        0        0 WD  RW   Failed Over
.
.
.
0010 0000   R2:3  RW RW NR   S..      212        0 RW  WD   Failed Over
0011 0001   R2:3  RW RW NR   S..       61        0 RW  WD   Failed Over
```

Note that there are invalid R1 tracks as activity continues on the R2 side.

Perform the following on the secondary system:

```
Symrdf -g prodrootvg swap –refresh R1
Symrdf -g prodrootvg establish
```

This command updates the R1 side with the changes made since the failover command, and then reverses the roles of R1/R2 disks. Both copies from this point forward would stay in a synchronized state.

Return the servers to production state

To return the servers to their normal state (local host on external `rootvg` and remote host on internal `rootvg`), reboot the remote host on its original internal `rootvg` copy.

```
>bootlist -m normal –o
hdisk19 # (This is output from the above command.)


>bootlist -m normal hdisk1

>bootlist -m normal -o
hdisk1  # (This is output from the above command.)

>Shutdown -Fr
```

When the remote host comes back on its internal disks, you can then set the SRDF pairs back to their normal state and reboot the local host with them. Use the appropriate command to return the R1/R2 pairs to the original state. If you chose to "failover", then run "failback." If you chose to "split", then run "establish."

```
>symrdf -g prodrootvg failback

Execute an RDF 'Failback' operation for device
group 'prodrootvg' (y/[n]) ? y

An RDF 'Failback' operation execution is
in progress for device group 'prodrootvg'. Please wait...

    Write Disable device(s) on RA at target (R2).............Done.
    Suspend RDF link(s).....................................Done.
    Merge device track tables between source and target.....Started.
    Devices: 0000-0001 ..............................Merge device
                      track tables between source and target.....Done.
    Resume RDF link(s)......................................Done.
    Read/Write Enable device(s) on SA at source (R1).......Done.

The RDF 'Failback' operation successfully executed for
device group 'prodrootvg'.
```

Verify that the R1 and R2 devices are synchronized and there are no invalid tracks.

```
>symrdf -g prodrootvg query

Device Group (DG) Name             : prodrootvg
DG's Type                          : RDF1
DG's Symmetrix ID                  : 000187900087


       Source (R1) View                Target (R2) View     MODES
--------------------------------    ------------------------ ----- ------------
            ST                  LI      ST
Standard     A                   N       A
Logical      T  R1 Inv   R2 Inv  K       T  R1 Inv   R2 Inv       RDF Pair
Device  Dev  E  Tracks   Tracks  S Dev   E  Tracks   Tracks MDA   STATE
-------------------------------- -- ------------------------ ----- ------------

DEV001  0000 RW       0        0 RW 0010 WD       0        0 S..   Synchronized
DEV002  0001 RW       0        0 RW 0011 WD       0        0 S..   Synchronized

Total          -------- --------           -------- --------
  Track(s)            0        0                  0        0
  MB(s)             0.0      0.0                0.0      0.0
```

You can boot the local host back on its internal disk to verify a problem was fixed, or use the SMS multiboot menus to change the bootlist to its external `rootvg` copy and reboot off the external devices. Use `lspv` to verify which copy is your active one:

```
>lspv

hdisk0          00601910967f88f3                    rootvg
hdisk1          00015458bfe2e31f                    rootvg
hdisk2          0002175f3043527f                    altinst_rootvg
hdisk3          0002175f30435da6                    altinst_rootvg

>bootlist -m normal -o
hdisk0 # (This is output from the above command.)

>bootlist -m normal hdisk2

>bootlist -m normal -o
hdisk2 # (This is output from the above command.)

>shutdown -Fr
```

Once your local host has been booted of external disk, you may have to once again run `symcfg discover` to undo the link changes.

```
>symrdf -g prodrootvg query

Error opening the gatekeeper device for communication to the Symmetrix
```

If this happens it is due to the paths on the remote host being severed. Run **symcfg discover** to dynamically repair.

```
>symcfg discover

This operation may take up to a few minutes. Please be patient...
```

Notice now that you have the R2 devices in a Defined state on the external disk copy of `rootvg`:

```
>lsdev -Cc disk

hdisk0  Available 10-60-00-8,0 16 Bit SCSI Disk Drive
hdisk1  Available 10-60-00-9,0 16 Bit SCSI Disk Drive
hdisk2  Available 20-58-01     EMC Symmetrix FCP RDF1 Raid1
hdisk3  Available 20-58-01     EMC Symmetrix FCP RDF1 Raid1
.
.
.
hdisk19 Defined   20-58-01     EMC Symmetrix FCP RDF2 Raid1
hdisk20 Defined   20-58-01     EMC Symmetrix FCP RDF2 Raid1
```

Troubleshooting

Because booting from an external Fibre Channel device is so similar to booting from an external SCSI device, many of the procedures and practices for troubleshooting already exist and can be applied to similar situations when troubleshooting Fibre Channel boot issues.

NVRAM corruption and SAN changes

If the NVRAM of the system becomes corrupted or if changes are made to the SAN environment that cause the WWPN, SCSI ID (destination ID), or LUN ID to change, you will need to redo some portions of this procedure to get booting working again.

The steps are:

1. After the banner screen and the word keyboard have appeared, press **F1** (**1** on a TTY) to enter the SMS menus.
2. Select the Multiboot option.

   After selecting Multiboot, the bus is scanned for any disks that contain the AIX boot image.
3. Select the boot device option. In this screen any bootable device found during the device scan can be selected. Selecting the boot device(s) order here will rewrite the bootlist, placing it into the order you have specified here.
4. Exit the menus to allow the system to boot. If after the scan no suitable boot image was found, check your cable connections.

   If the system continues to fail to boot, continue to "Open firmware," discussed next, to collect some of the adapter settings directly.

Open firmware

Because the adapter topology is auto-discovered during system bootup, the open firmware prompt is usually used only to verify what NVRAM contains. The primary setting of concern in NVRAM is the topology. If the topology is set incorrectly, or for some reason is not auto-detected, it can be set manually. In the event that NVRAM has the incorrect topology, you can use the following section to set this manually. If you have manually set the topology and your adapter still has problems logging in to the SAN, you should verify that the firmware has upgraded successfully and that your adapter is not defective.

Locating the Fibre Channel adapter

To locate the Fibre Channel adapter:

1. Boot the system to an Open Firmware prompt.
2. Press **F8** (**8** on a TTY to serial port) after the RS/6000 banner appears and after the word keyboard appears, or select the **`OK`** prompt from the Multiboot menu if multiboot startup is enabled.
3. Use the command **`dev / ls`** to display the device tree for the system (use **`CTRL-S`** and **`CTRL-Q`** to stop and resume scrolling).

The output is similar to the following:

```
00cca750:       /pci@f8700000
00d34c00:             /fibre-channel@b
00d41540:                   /disk
00ccd5a0:       /pci@f8c00000
00d42488:             /fibre-channel@b
00d4e538:                   /disk
00cd03f0:       /pci@f8d00000
00d4f450:             /token-ring@c
00cd3240:       /pci@f8e00000
00d52db8:             /IBM, longbow@f
00d54870:             /IBM. longbow@f,1
00cd6090:       /pci@f8f00000
00d58820:             /fibre-channel@c
00d648b0:                   /disk
00cd8ee0:       /reserved@f8000000
```

Note: The bootable Fibre Channel adapters will have **fibre-channel** in their name. In the above example, `/pci@f8700000/fibre-channel@b` is a Fibre Channel adapter that is a child of the `/pci@f8700000` device. If there is more than one Fibre Channel adapter in the system, you can determine the mapping between devices and physical slot location by looking at the properties of each device.

4. Before you can act on a Fibre Channel adapter, you must first tell open firmware which adapter to execute commands against by selecting it. As in the example in step 3, to select our first Fibre Channel adapter visible we use the **`select-dev`** command:

   ```
   " /pci@f8700000/fibre-channel@b" select-dev
   ```

   Note: There is a space between the first double quote and the slash, and the double quotes are required for the command to work properly.

5. Show the properties of the device with the **`.properties`** command. Included among this information is the `ibm,loc-code` property that will have a format such as `U0.1-P1-I8/Q1`. This indicates the adapter is in slot 8 (from I8) of drawer 0 (from .1, I/O drawers are numbered starting from zero, but the location codes start at one). Refer to your machine's installation and service guide for more information on location codes.
6. Repeat these steps for any other adapters to determine their corresponding physical locations.

Manually setting the Fibre Channel adapter topology

To manually set the topology:

1. Once the adapter has been selected using `select-dev`, you must use the **.topology** command to determine the current topology:

   ```
   .topology
   fc-al
   ```

2. If the topology is not set to the proper configuration, you can set it manually by using the following sets of commands. These commands are specific to the adapter selected so setting the topology for the selected adapter will not change the topology for others:

   `Set-ptp` (for connecting to a switch)

   `Set-fc-al` (for connecting to public or private loop)

3. Unselect the adapter by entering **`unselect-dev`**.

4. If you have more adapters, repeat the selection process starting with step 1; otherwise, return to SMS to select your boot device and add it to the bootlist.

   If the adapter still does not log in as expected after setting the topology manually, you should look at the configuration settings of your environment to ensure they have been configured properly or try a new host bus adapter.

CHAPTER 2

Virtual I/O Server

This chapter contains the following information:

Note: Any general references to Symmetrix or Symmetrix array include the VMAX3 Family, VMAX, and DMX.

Virtual I/O Server overview

The Virtual I/O Server (VIOS) provides virtual storage and shared Ethernet capability to client logical partitions on a System p-based system. It allows a physical adapter with attached disks on the VIOS partition to be shared by one or more Virtual I/O Client (VIOC) partitions, enabling clients to consolidate and minimize the number of physical adapters.

The VIOS is created using the Hardware Management Console (HMC). The VIOS installation media ships with the System p system that is configured with the Advanced POWER Virtualization feature and can be installed by:

- Media (assigning the DVD-ROM drive to the partition and booting from the media).
- The HMC (inserting the media in the DVD-ROM drive on the HMC and using the **`installios`** command).
- Using NIM.

The VIOS supports AIX 5.3, 6.1, and pLinux, as Virtual I/O clients.

The VIOS provides the Virtual SCSI server and Shared Ethernet Adapter virtual I/O function to AIX client partitions. This VIOS partition is not intended to run applications at this time or for general user login but specific commands, such as **oem_setup_env**, allow the installation of the software products. These include Dell EMC ODM Support Software and PowerPath multipathing software.

POWER Hypervisor and virtual I/O

With the introduction of micropartitioning, the ability to dedicate physical hardware adapter to each partition becomes impractical. Virtualization of I/O devices allows many partitions to communicate with each other, and access networks and storage devices external to the server, without dedicating physical I/O resources to an individual partition. Many of the I/O virtualization capabilities introduced with the System p server products are accomplished by functions designed into the POWER Hypervisor.

The POWER Hypervisor does not own any physical I/O devices, and it does not provide virtual interfaces to them. All physical I/O devices in the system are owned by logical partitions. Virtual I/O devices are owned by the VIOS, which provides access to the real hardware that the virtual device is based on.

The POWER Hypervisor implements the following operations required by system partitions to support virtual I/O:

- Provides control and configuration structures for virtual adapter images required by the logical partitions.
- Operations that allow partitions controlled and secure access to physical I/O adapters in a different partition.

The POWER Hypervisor allows for the virtualization of I/O interrupts. To maintain partition isolation, the POWER Hypervisor controls the hardware interrupt management facilities. Each logical partition is provided controlled access to the interrupt management facilities using hcalls. Virtual I/O adapters and real I/O adapters use the same set of hcall interfaces.

Supported I/O types

Three types of virtual I/O adapters are supported by the POWER Hypervisor:

- SCSI
- Ethernet
- Serial console (virtual TTY)

Virtual I/O adapters are defined by system administrators during logical partition definition. Configuration information for the virtual adapters is presented to the partition operating system by the system firmware.

Virtual SCSI support

The System p server uses SCSI as the mechanism for virtual storage devices. This is accomplished using two paired adapters: a virtual SCSI server adapter and a virtual SCSI client adapter. These adapters are used to transfer SCSI commands between partitions. To the operating system, the virtual SCSI client adapter is no different than a typical SCSI adapter. The SCSI server, or target, adapter is responsible for executing any SCSI command it receives. It is owned by the Virtual I/O Server partition.

Virtual SCSI operation is dependent on two system functions implemented in the POWER Hypervisor:

- Message queuing function that enables the passing of messages between partitions, with an interrupt mechanism to confirm receipt of the message.
- DMA function between partitions that provides control structures for secure transfers between logical partition memory spaces.

Virtual Ethernet switch support

The POWER Hypervisor provides a Virtual Ethernet switch function that allows partitions on the same server a means for fast and secure communication. It is based on the IEEE 802.1Q VLAN standard. No physical I/O adapter is required when creating a VLAN connection between partitions, and no access to an outside network is required. Each partition operating system sees the VLAN switch as an Ethernet adapter, without the physical link properties and asynchronous data transmit operations.

Virtual TTY console support

Each partition needs to have access to a system console. Tasks such as operating system install, network setup, and some problem analysis activities require a dedicated system console. The POWER Hypervisor provides virtual console using a virtual TTY or serial adapter and a set of hypervisor calls to operate on them.

Virtual I/O Server configuration

The Virtual I/O Server (VIOS) provides the virtual SCSI and shared Ethernet adapter virtual I/O to client partitions. This is accomplished by assigning physical devices to the VIOS partition, then configuring virtual adapters on the clients to allow communication between the client and the VIOS.

Using virtual I/O devices facilitates the following functions:

- Sharing of physical resources between partitions on a System p system.
- Providing Virtual SCSI and Shared Ethernet Adapter function to client partitions.
- Creation of partitions without requiring additional physical I/O resources.
- Creation of more partitions than I/O slots or physical devices with the ability for partitions to have dedicated I/O, virtual I/O, or both.
- Maximizing the utilization of physical resources on a System p system.

Virtual SCSI disk

Virtual SCSI facilitates the sharing of physical disk resources (I/O adapters and devices) between logical partitions. Virtual SCSI enables partitions to access SCSI disk devices without requiring that physical resources be allocated to the partition. Partitions maintain a client/server relationship in the Virtual SCSI environment. Partitions that contain Virtual SCSI devices are referred to as client partitions while the partition that owns the physical resources (adapters, devices) is the VIOS.

It is preferred that the Virtual SCSI disks are defined as physical volumes, rather than logical volumes in the VIOS. All standard conventional rules apply to the physical volumes. The physical volumes appear as real devices (hdisks) in the client partitions and can be used as a boot device and as a NIM target. Once a virtual disk is assigned to a client partition, the VIOS must be available before the client partitions are able to boot.

Figure 3 on page 88 shows a standard virtual SCSI configuration.

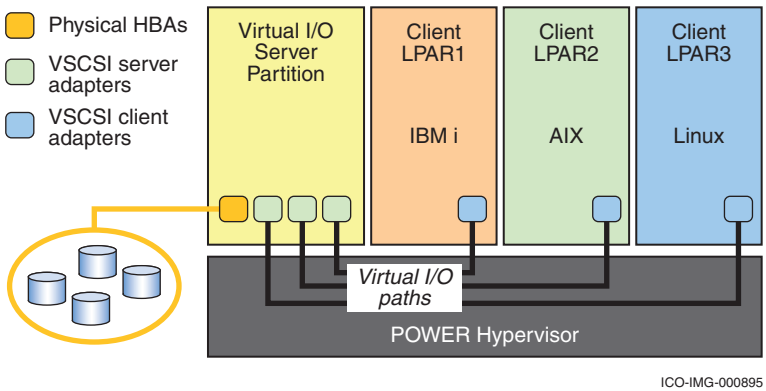


Figure 3 Virtual SCSI configuration example

Basic configuration

This section discusses the basic configuration scenarios of the Virtual I/O Server (VIOS) and virtual client partitions. It is assumed that the system administrator is familiar with the installation procedure of the Virtual I/O Server and Virtual I/O Client Software. Additional information can be obtained from the IBM Advanced POWER Virtualization on IBM eServer p5 Servers Redbook:SG24-7940-00:

```
http://www.redbooks.ibm.com/redpieces/abstracts/sg247940.html
```

Creating the virtual target device on the Virtual I/O Server

The basic command to map the Virtual SCSI with the physical hdisk is:

```
mkvdev -vdev TargetDevice -vadapter VirtualSCSIServerAdapter
```

Type the **`lsdev -virtual`** command to make sure that your new virtual SCSI adapter is available:

```
$ lsdev -virtual
```

Output example:

```
name         status       description
vhost0       Available    Virtual SCSI Server Adapter
vhost1       Available    Virtual SCSI Server Adapter
vhost2       Available    Virtual SCSI Server Adapter
vtscsi0      Available    Virtual Target Device - Disk
```

The **`lsmap -all`** command shows the logical connections between the newly created devices:

```
$ lsmap -vadapter vhost0
```

Output example:

```
SVSA               Physloc                     Client Partitiion
--------------- ----------------------------------
vhost0             U9111.520.10DDEEC-V1-C20   0x00000002
VTD                vtscsi0
LUN                0x8100000000000000
Backing device     hdisk2
Physloc            U787B.001.DNW0F92-P1-C2-T1-L1
```

Here you also see the physical location being a combination of the slot number (20 in the example) and the logical partition ID.

You can now activate your partition into the SMS menu and install the AIX operating system onto the virtual disk or add an additional virtual disk using the **`mkvdev`** or **`cfgmgr`** command.

Advanced configuration

This section discusses the different configuration scenarios of the Virtual I/O Server (VIOS) and Virtual I/O Client (VIOC) partitions with EMC-attached storage.

Creating virtual I/O configurations using Dell EMC-attached storage

The following is a guideline for integrating Dell EMC-attached storage into a Virtual I/O configuration. The following are the minimum requirements:

- System p server with appropriate Virtual I/O feature enabled.
- Virtual I/O Server must have minimum FixPack 6.2 (includes APAR IY58231), plus efix IY71303.062905.epkg.Z.

  Note: Consult the Dell EMC Simple Support Matrices for the most up-to-date information and for additional minimum requirements.

- VIOC LPARs must be minimum AIX 5.3 ML 2 (APAR IY69190), with APARs IY70082.

  Note: Consult the Dell EMC Simple Support Matrices for the most up-to-date information and for additional minimum requirements.

- EMC subsystems:
  - EMC Symmetrix DMX, Symmetrix, and VMAX.
  - VNX Series
  - CLARiiON CX, CX3, and CX4
- As is the case for non-Virtual I/O environments, EMC ODM Support Software package is required. The ODM package is implemented in the partition (VIOS or VIOC) to which devices are physically attached. Obtain the minimum ODM package version 5.2.0.3 from the Dell EMC FTP site:

  `ftp://ftp.emc.com/pub/elab/aix/ODM_DEFINITIONS`

- Multipathing requirements:

  Multipathing is required and is supported in the following configurations:

  - PowerPath 4.4.2.2 or above is supported in:
    - VMAX/Symmetrix
    - VNX series
    - CLARiiON
  - IBM Native MPIO is supported in:
    - VMAX/Symmetrix
    - VNX series
    - CLARiiON
      - AIX 5.3 TL8 SP4
      - AIX 5.3 TL9
      - AIX 6.1TL1 APAR IZ32812
      - AIX 6.1TL2
      - VIOS 1.5.2.0 SP4
      - VIOS 2.1
    - CLARiiON in ALUA configurations
      - AIX 5.3 TL8 SP8
      - AIX 5.3 TL9 SP5
      - AIX 5.3 TL10 SP2
      - AIX 5.3 TL11
      - AIX 6.1 TL1 SP6
      - AIX 6.1 TL2 SP5
      - AIX 6.1 TL3 SP2
      - AIX 6.1 TL4

– VIOS 2.1 FP22.1

- In a dual VIO environment where LUNs from a CLARiiON or VNX array is presented as VDASD devices to VIO clients, if the VIO servers are registered with failover mode 1 (PNR) and are using native MPIO to manage the multipathing, the hcheck_cmd attribute for the shared VDASD devices on the VIO client needs to be changed from the default value of test_unit_rdy to inquiry. This is to prevent random path failures for the shared CLARiiON or VNX hdisks on the VIO servers. See EMC Knowledgebase solution emc210546 for more information.

  Note: This step is not required if `failovermode 4` (ALUA) is used on the VIO servers.

- PowerPath capabilities and configuration guidance when running in a VIOS LPAR are the same as for running PowerPath under AIX 5.3.
- PowerPath is allowed and utilizable in a VIOC LPAR for paths and devices which are physically attached to the VIOC.
- The System p system can have multiple VIOS LPARs providing access to the same LUN for the same VIOC for Standard Device configurations (no multipathing) and multipathing devices under the control of PowerPath and MPIO. For Standard Device and PowerPath configurations, the Virtual I/O Server device reservation setting **`reserve_lock`** must be changed (via the **`chdev`** command) to **`no`**. For MPIO configurations and PowerPath v5.5 adn later, the Virtual I/O Server device reservation setting **`reserve_policy`** must be changed (via the **`chdev`** command) to **`no_reserve`**.
- VIOC with Dell EMC-supported devices attached, either directly or through Virtual I/O Server, can be booted from an exported internal SCSI device. A VIOC can be SAN booted from a Dell EMC-supported device, either accessed via the VIOS or directly attached to the VIOC. Configurations in which multiple VIOSs are exporting a given `rootvg` LUN to the same VIOC are supported at minimum IOS level 1.3.
- Logical Partition Mobility (LPM)

  Logical Partition Mobility is supported on Power6 and Power7-based servers:

  - PowerPath 5.1 and higher
    – Symmetrix
    – VNX series
    – CLARiiON
  - IBM Native MPIO
    – Symmetrix
    – VNX series
    – CLARiiON

      – AIX 5.3 TL8 SP4
      – AIX 5.3 TL9
      – AIX 6.1TL1 APAR IZ32812
      – AIX 6.1TL2
      – VIOS 1.5.2.0 SP4
      – VIOS 2.1

    – CLARiiON in ALUA configurations
      – AIX 5.3 TL8 SP8

  - AIX 5.3 TL9 SP5
  - AIX 5.3 TL10 SP2
  - AIX 5.3 TL11
  - AIX 6.1 TL1 SP6
  - AIX 6.1 TL2 SP5
  - AIX 6.1 TL3 SP2
  - AIX 6.1 TL4
  - VIOS 2.1 FP22.1
- HMC Version V7R3.3.0.1 and higher
- VIOS 1.5.2.1 and higher

- NPIV support
  - NPIV is supported with minimum VIOS 2.1 for
    - Symmetrix
    - VNX series
    - CLARiiON

Note: NPIV and LPM support requires device **`reserve_policy=no_reserve`** or **`reserve_lock=no`** for all SAN devices connected to the VIOC.

Figure 4 shows a managed system configured to use NPIV.

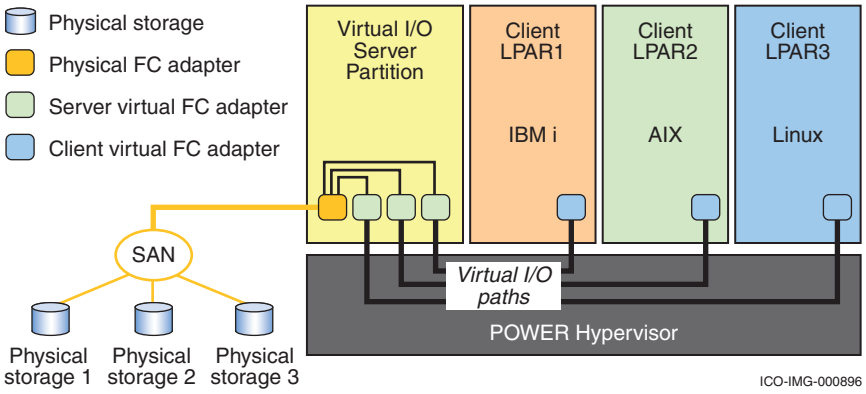


Figure 4 Virtual Fibre Channel architecture example

Important Virtual I/O information

Special attention should be noted when planning Virtual I/O Server, Virtual I/O Client installations, and migration requirements that are not specifically storage-centric (this is general guidance and not unique to Dell EMC). There are different methods to identify uniquely a disk for use in Virtual I/O Server, such as:

- Unique device identifier (UDID)
- Physical volume identifier (PVID)
- IEEE volume identifier

Each of these methods may result in different data formats on the disk. This implies device block offset differences. The preferred disk identification method for virtual disks is the use of UDIDs. An example of this use is minimum PowerPath version 4.4.2.2 and MPIO.

Most non-UDID disk storage multipathing software products use the PVID method instead of the UDID method. Because of the different data formats associated with this, customers with non-UDID environments should be aware that certain future actions performed in the VIOS

LPAR may require data migration, that is, some type of backup and restore of the attached disks. When used with PVID method, the traditional BCV, Dell EMC SRDF™, Dell EMC MirrorView™, and ADMsnap replication copies are not supported as restoration methods to exported devices.

The following configurations and conversions adhere to this data format change provided the Virtual Target Device and Virtual SCSI Server Adapter ODM instances become non-persistent. A non-persistent example would include existing exported physical hdisks as virtual devices, using the PVID method, which are removed with the **`rmdev -dev`** command with the **`recursive`** flag (**`rmdev -dev vhost0 -recursive`**).

These actions may include, but are not limited to:

- An existing Virtual I/O configuration with Non-MPIO device (no multipathing) configuration is converted to a MPIO configuration.
- An existing Virtual I/O configuration with MPIO device configuration is converted to a Non-MPIO Standard device (no multipathing) configuration.
- An existing Virtual I/O Configuration which is converted from the PVID to the UDID method of disk identification.
- An existing Virtual I/O Configuration which is converted from the UDID to the PVID method of disk identification.
- An existing Virtual I/O configuration, previously approved via EMC RPQ, with PowerPath 4.4, 4.4.2 which is converted to PowerPath version 4.4.2.2 and higher. Refer to "Virtual I/O Server and client device migration" on page 99 for more information.
- Possible future enhancements to VIO.

Virtual I/O Server software installation

This section provides a description of the installation procedure of the Virtual I/O Server (VIOS) software to the formerly created partition VIO_Server_1.

Installation of the VIOS partition is performed from a special **mksysb** CD that is provided to customers that order the Advanced POWER Virtualization feature, at an additional charge. This is dedicated software only for the VIOS operations, so the VIOS software is only supported in VIOS partitions. VIOS partitions are not intended to run applications at this time. An exception to this is EMC ODM software package and PowerPath multipathing software.

The following steps show the installation using the CD install device:

1. To activate the VIO_Server_1 partition, right-click the partition name and select the **Activate** bar.
2. Select the default profile you used to create this server. Select the default profile, activate the **Open** a terminal window or console session checkbox, and click **Advanced**.
3. Choose the SMS boot mode, and click **OK** to return to the previous window.

   In the previous window, click **OK** to activate the partition and launch a terminal window.
4. Proceed through the installation procedure as for any other AIX installation, choosing CD as the installation device.
5. When the installation procedure has finished, use **padmin** for **username** at the login prompt, choose a new password, and accept the license using the **license -accept** command at the $ prompt.

   You can use the **lspv** command to show the available disks.

   With this step you have finished the installation of the Virtual I/O Server. The **VIO_Server_1** partition is now ready for further configuration.

Virtual I/O Server installation with Dell EMC-attached storage

The following steps show the installation of Dell EMC-specific software:

1. Log into the VIO_Server_1 partition and use the **oem_setup_env** command to access out of the restricted shell.

   Type the **lspv** command to display your configured devices

   > **lspv**

   Output example:

   ```
   hdisk0      00cfa11d15552dd5   rootvg          active
   ```

2. Download ODM support software version 5.2.0.3 or higher from the EMC FTP server and install the appropriate filesets.

   In this particular configuration we will be installing the EMC Symmetrix AIX Support Software, Symmetrix Fibre Channel software, VNX series Fibre Channel software, and CLARiiON Fibre Channel software.

3. Type the **exit** command to exit the shell. Then type the `shutdown -restart` command to reboot the server.

4. Log in to the VIO_Server_1 partition and type the **oem_setup_env** command to access out of the restricted shell.

   Use the **lsdev -Cc** disk command to display your configured devices.

   > **lsdev -Cc disk**

   Output example:

   ```
   hdisk0 Available 05-08-00-3,016 Bit LVD SCSI Disk Drive
   ```

5. Set up the appropriate device zoning to configure the rest of your EMC storage devices.

   Once this is complete, type the **emc_cfgmgr** command to configure your devices.

   Type the **lsdev -Cc** disk command to display your configured devices.

   ```
   > /usr/lpp/EMC/Symmetrix/bin/emc_cfgmgr
   > lsdev -Cc disk
   ```

   Output example:

```
hdisk0 Available 05-08-00-3,0    16 Bit LVD SCSI Disk Drive
hdisk1 Available 06-08-02        EMC Symmetrix FCP Raid1
hdisk2 Available 06-08-02        EMC Symmetrix FCP Raid1
hdisk3 Available 03-08-02        EMC Symmetrix FCP Raid1
hdisk4 Available 03-08-02        EMC Symmetrix FCP Raid1
hdisk5 Available 06-08-02        EMC CLARiiON FCP RAID 1/0 Disk
hdisk6 Available 06-08-02        EMC CLARiiON FCP RAID 1/0 Disk
hdisk7 Available 03-08-02        EMC CLARiiON FCP RAID 1/0 Disk
hdisk8 Available 03-08-02        EMC CLARiiON FCP RAID 1/0 Disk
```

6. Install PowerPath version 4.4.2.2 or higher.

   Note: Version 4.4.2.2 requires a previous installation of version 4.4 and 4.4.2.

   Do not type the **powermt config** command after these versions are installed.

Once PowerPath is installed, type the **powermt config** command to configure your PowerPath devices.

Type the **lsdev -Cc disk** command to display your configured devices:

> **powermt config**

> **lsdev -Cc disk**

Output example:

```
hdisk0       Available 05-08-00-3,0     16 Bit LVD SCSI Disk Drive
hdisk1       Available 06-08-02         EMC Symmetrix FCP Raid1
hdisk2       Available 06-08-02         EMC Symmetrix FCP Raid1
hdisk3       Available 03-08-02         EMC Symmetrix FCP Raid1
hdisk4       Available 03-08-02         EMC Symmetrix FCP Raid1
hdisk5       Available 06-08-02         EMC CLARiiON FCP RAID 1/0 Disk
hdisk6       Available 06-08-02         EMC CLARiiON FCP RAID 1/0 Disk
hdisk7       Available 03-08-02         EMC CLARiiON FCP RAID 1/0 Disk
hdisk8       Available 03-08-02         EMC CLARiiON FCP RAID 1/0 Disk
hdiskpower0  Available 06-08-02         PowerPath Device
hdiskpower1  Available 06-08-02         PowerPath Device
hdiskpower2  Available 06-08-02         PowerPath Device
hdiskpower3  Available 06-08-02         PowerPath Device
hdiskpower4  Available 06-08-02         PowerPath Device
hdiskpower5  Available 06-08-02         PowerPath Device
hdiskpower6  Available 06-08-02         PowerPath Device
hdiskpower7  Available 06-08-02         PowerPath Device
```

Your Virtual I/O Server setup is now complete, and you can now proceed in configuring your Virtual I/O Client partition.

Virtual I/O client device configuration with Dell EMC-attached storage

This section shows the steps required to configure your virtual SCSI devices:

1. Before we map the Virtual SCSI with the physical hdisk, we will first place PVID on the `hdiskpower` devices. This will assist in understanding the device relationships between the Virtual I/O Server (VIOS) and the Virtual I/O Client (VIOC).

   Type the following command (where <*x*> is the device instance) on the VIOS to the configured `hdiskpower` devices:

   > **chdev -l hdiskpower**<*x*> **-a pv=yes**

2. Type the **lspv** command to display your configured devices:

   > **lspv**

   Output example:

```
hdisk0          00cfa11d15552dd5       rootvg     active
hdisk1          none                   None
hdisk2          none                   None
hdisk3          none                   None
hdisk4          none                   None
hdisk5          none                   None
hdisk6          none                   None
hdisk7          none                   None
hdisk8          none                   None
hdiskpower0     00cfa11d72d3aaa3       None
hdiskpower1     00cfa11d72d77343       None
hdiskpower2     00cfa11d72dde22c       None
hdiskpower3     00cfa11d72d53f81       None
```

```
hdiskpower4     00cfa11d72dba7e8       None
hdiskpower5     00cfa11d72dea14a       None
hdiskpower6     00cfa11d72dd8fdd       None
hdiskpower7     00cfa11d72d6eda6       None
```

3. Type the **exit** command to exit the shell.

   Type the **lsdev -virtual** command to confirm that your virtual SCSI adapter is available (as identified by the:

   ```
   $ lsdev -virtual
   ```

   Output example:

```
name         status       description
ent2         Available    Virtual I/O Ethernet Adapter (l-lan)
vhost0       Available    Virtual SCSI Server Adapter
vhost1       Available    Virtual SCSI Server Adapter
```

4. Create a virtual target device, which maps the Virtual SCSI Server adapter vhost0 to the `hdiskpower` device. When you do not use the `-dev` flag, the default name of the Virtual Target Device adapter is **vtscsix**. This will be your client target boot device.

   Type the **mkvdev** command for each boot device instance:

   ```
   $ mkvdev -vdev hdiskpower0 -vadapter vhost0
   ```

   Output example:

   ```
   vtscsi0Available
   ```

5. Type the **mkvdev** command for each additional device instance using an alternate vhostx instance:

   Example:

   ```
   $ mkvdev -vdev hdiskpower1 -vadapter vhost1
   ```

   ```
   vtscsi1  Available
   ```

6. Type the **lsdev** command to show the newly created Virtual Target Device:

   ```
   $ lsdev -virtual
   ```

   Output example:

```
name         status       description
ent2         Available    Virtual I/O Ethernet Adapter (l-lan)
vhost0       Available    Virtual SCSI Server Adapter
vhost1       Available    Virtual SCSI Server Adapter
vtscsi0      Available    Virtual Target Device - Disk
vtscsi1      Available    Virtual Target Device - Disk
vtscsi2      Available    Virtual Target Device - Disk
vtscsi3      Available    Virtual Target Device - Disk
vtscsi4      Available    Virtual Target Device - Disk
vtscsi5      Available    Virtual Target Device - Disk
vtscsi6      Available    Virtual Target Device - Disk
vtscsi7      Available    Virtual Target Device - Disk
```

7. Type the **lsmap** command to show the logical connections between the newly created devices:

   ```
   $ lsmap -all
   ```
   Output example:

```
SVSA            Physloc                          Client PartitionID
--------------------------------------------------
vhost0          U9111.520.10DDEEC-V1-C20      0x00000002
VTD             vtscsi0
LUN             0x8100000000000000
Backing device  hdiskpower0
Physloc         U787B.001.DNW0F92-P1-C2-T1-L0

vhost1          U9111.520.10DDEEC-V1-C21      0x00000002
VTD             vtscsi1
LUN             0x8200000000000000
Backing device  hdiskpower1
Physloc         U787B.001.DNW0F92-P1-C2-T1-L1

VTD             vtscsi2
LUN             0x8300000000000000
Backing device  hdiskpower2
Physloc         U787B.001.DNW0F92-P1-C2-T1-L2

VTD             vtscsi3
LUN             0x8400000000000000
Backing device  hdiskpower3
Physloc         U787B.001.DNW0F92-P1-C2-T1-L3

VTD             vtscsi4
LUN             0x8500000000000000
Backing device  hdiskpower4
Physloc         U787B.001.DNW0F92-P1-C2-T1-L4

VTD             vtscsi5
LUN             0x8600000000000000
Backing device  hdiskpower5
Physloc         U787B.001.DNW0F92-P1-C2-T1-L5

VTD             vtscsi6
LUN             0x8700000000000000
Backing device  hdiskpower6
Physloc         U787B.001.DNW0F92-P1-C2-T1-L6

VTD             vtscsi7
LUN             0x8800000000000000
Backing device  hdiskpower7
Physloc         U787B.001.DNW0F92-P1-C2-T1-L7
```

8. Now that the Virtual I/O Client devices are exported, we can proceed in activating the Client Partition ID into the SMS menu and install the AIX operating system on virtual target device `vtscsi0` (`vtscsi0`, LUN 0x81, `hdiskpower0`).

The installation procedures are complete once the AIX operating system is installed on the virtual target device.

Virtual I/O Server and client device migration

This section is a guideline only for devices that have previously been configured to a Virtual I/O Server (VIOS) and Virtual I/O Client (VIOC) partition.

There have been several Dell EMC-approved RPQs (pre-GA) that support Dell EMC-attached storage in a Virtual I/O configuration environment. These include Symmetrix with MPIO and Symmetrix, using PowerPath version 4.4 or 4.4.2. The following steps act as a guideline for migrating into a PowerPath 4.4.2.2 or higher GA supported environment.

VIOS uses several methods to uniquely identify a disk for use in as a virtual SCSI disk:

- Unique device identifier (UDID)
- Physical volume identifier (PVID)
- IEEE volume identifier

Each of these methods can result in different data formats on the disk. The preferred disk identification method for virtual disks is the use of UDIDs. Currently, MPIO and PowerPath 4.4.2.2 and higher use the UDID method.

IMPORTANT

Dell EMC-exported physical hdisk devices configured with no multipathing software products and PowerPath versions prior to 4.4.2.2 use the PVID method instead of the UDID method. Customers with non-UDID environments should be aware that certain future actions performed in the VIOS LPAR may require data migration, that is, some type of backup and restore of the attached disks.

Actions that may require data migration include, but are not limited to:

- Conversion from a Non-MPIO environment to MPIO
- Conversion from the PVID to the UDID method of disk identification
- Removal and rediscovery of the Disk Storage ODM entries
- Updating non-MPIO multipathing software under certain circumstances
- Possible future enhancements to VIO

Supported migration environments include:

- MPIO (UDID) to PowerPath (UDID) 4.4.2.2 and higher
- PowerPath 4.4.2.2 (UDID) and higher to MPIO (UDID)
- PowerPath (PVID) 4.4, 4.4.2 to PowerPath (UDID) 4.4.2.2 or higher

Unsupported migration environments include:

- MPIO (UDID) to non-multipathing (PVID)
- Non-multipathing (PVID) to MPIO (UDID)
- PowerPath (PVID) 4.4, 4.4.2 to MPIO (UDID)
- MPIO (UDID) to PowerPath (PVID) 4.4 or 4.4.2
- PowerPath (UDID) 4.4.2.2 and higher to non-multipathing (PVID)
- Non-Multipathing (PVID) to PowerPath (UDID) 4.4.2.2 and higher

Symmetrix MPIO (UDID) to PowerPath (UDID) device migration

The following procedure outlines the steps required for migrating a previously configured Symmetrix MPIO (UDID) device to a PowerPath (UDID) device. It is highly recommended that all critical data be backed up before proceeding with these following steps.

This configuration consists of a Virtual I/O Server with EMC Symmetrix devices which are under the control of MPIO. These devices are also exported to the Virtual I/O client as virtual scsi devices.

1. Log in to the Virtual I/O Client and verify all volume groups are varied off. Record the PVID that resides on the volume groups by using the **`lspv`** command. In this particular case `rootvg` is 00cfa11d72d3aaa3:

   > **`lspv`**

   Output example:

```
hdisk0       00cfa11d72d3aaa3       rootvg    active
hdisk1       00cfa11d72d77343       vg1
hdisk2       00cfa11d72dde22c       vg2
hdisk3       00cfa11d72d53f81       vg3
```

2. Type the **`lsdev -Cc disk`** command to display your configured devices. Once this is complete, shut down the client partition using the **shutdown** command:

   Input and output example:

   > **`lsdev -Cc disk`**

```
hdisk0   Available    Virtual SCSI Disk Drive
hdisk1   Available    Virtual SCSI Disk Drive
hdisk2   Available    Virtual SCSI Disk Drive
hdisk3   Available    Virtual SCSI Disk Drive
```

   > **`shutdown`**

3. Log in to the Virtual I/O Server partition and type the **oem_setup_env** command to access out of the restricted shell.

   Then type the **`lsdev -Cc disk`** command to record your configured devices:

   > **`lsdev -Cc disk`**

   Output example:

```
hdisk0   Available 05-08-00-3,0     16 Bit LVD SCSI Disk Drive
hdisk1   Available 06-08-02         EMC Symmetrix FCP MPIO Raid1
hdisk2   Available 06-08-02         EMC Symmetrix FCP MPIO Raid1
hdisk3   Available 06-08-02         EMC Symmetrix FCP MPIO Raid1
hdisk4   Available 06-08-02         EMC Symmetrix FCP MPIO Raid1
```

4. Type the **lspv** command to record your configured devices:

   > **lspv**

   Output example:

```
hdisk0      00cfa11d15552dd5       rootvg       active
hdisk1      00cfa11d72d3aaa3       None
hdisk2      00cfa11d72d77343       None
hdisk3      00cfa11d72dde22c       None
hdisk4      00cfa11d72d53f81       None
```

5. Type the **exit** command to exit the shell.

   Then type the **lsdev -virtual** command to display your configured exported MPIO devices:

   $ **lsdev -virtual**

   Output example:

```
name     status       description
ent2     Available    Virtual I/O Ethernet Adapter (l-lan)
vhost0   Available    Virtual SCSI Server Adapter
vhost1   Available    Virtual SCSI Server Adapter
vtscsi0  Available    Virtual Target Device - Disk
vtscsi1  Available    Virtual Target Device - Disk
vtscsi2  Available    Virtual Target Device - Disk
vtscsi3  Available    Virtual Target Device - Disk
```

6. Type the **lsmap** command to record the logical connections between the virtual target devices and the backing devices. This information will be used when reconfiguring with `hdiskpower` devices:

   $ **lsmap -all**

   Output example:

```
SVSA            Physloc                          Client PartitionID
--------------------------------------------------------------------
vhost0          U9111.520.10DDEEC-V1-C20      0x00000002
VTD             vtscsi0
LUN             0x8100000000000000
Backing device  hdisk1
Physloc         U787B.001.DNW0F92-P1-C2-T1-L0

vhost1          U9111.520.10DDEEC-V1-C21      0x00000002
VTD             vtscsi1
LUN             0x8200000000000000
Backing device  hdisk2
Physloc         U787B.001.DNW0F92-P1-C2-T1-L1

VTD             vtscsi2
LUN             0x8300000000000000
Backing device  hdisk3
Physloc         U787B.001.DNW0F92-P1-C2-T1-L2

VTD             vtscsi3
LUN             0x8400000000000000
Backing device  hdisk4
Physloc         U787B.001.DNW0F92-P1-C2-T1-L3
```

7. Remove the configured exported devices using the **rmdev** command. This will unmap the logical connection between the virtual target devices and the backing devices. Perform this command to each vtscsi instance that is configured to the client partition.

   Input and output example:

   ```
   $ rmdev -dev vtscsi0
   vtscsi0 deleted
   $ rmdev -dev vtscsi1
   vtscsi1 deleted
   $ rmdev -dev vtscsi2
   vtscsi2 deleted
   $ rmdev -dev vtscsi3
   vtscsi3 deleted
   ```

8. Type the **oem_setup_env** command to access out of the restricted shell.

   Then use the **rmdev** command to remove your configured MPIO devices:

   ```
   > rmdev -dl hdisk1
   > rmdev -dl hdisk2
   > rmdev -dl hdisk3
   > rmdev -dl hdisk4
   ```

9. Type the following command to remove the Symmetrix Fibre Channel MPIO fileset:

   ```
   > installp -u EMC.Symmetrix.fcp.MPIO.rte
   ```

10. Download ODM support software version 5.2.0.3 or higher from the Dell EMC FTP server and install the appropriate filesets.

    In this particular configuration we will be installing the Symmetrix AIX Support Software and Symmetrix Fibre Channel Support Software.

11. Type the **emc_cfgmgr** command to configure the devices:

    ```
    > /usr/lpp/install/Symmetrix/bin/emc_cfgmgr
    ```

12. Install PowerPath software version 4.4.2.2 or higher.

    Note: Version 4.4.2.2 requires a previous installation of v4.4 and 4.4.2.

    Do not type the **powermt config** command after installing these versions.

    Once PowerPath is installed, type the **powermt config** command to configure your PowerPath devices. Then type the **lsdev -Cc disk** command to display your configured devices:

    ```
    > powermt config
    > lsdev -Cc disk
    ```

    Output example:

```
hdisk0          Available 05-08-00-3,0 16 Bit LVD SCSI Disk Drive
hdisk1          Available 06-08-02 EMC Symmetrix FCP Raid1
hdisk2          Available 06-08-02 EMC Symmetrix FCP Raid1
hdisk3          Available 06-08-02 EMC Symmetrix FCP Raid1
hdisk4          Available 06-08-02 EMC Symmetrix FCP Raid1
hdisk5          Available 03-08-02 EMC Symmetrix FCP Raid1
hdisk6          Available 03-08-02 EMC Symmetrix FCP Raid1
hdisk7          Available 03-08-02 EMC Symmetrix FCP Raid1
hdisk8          Available 03-08-02 EMC Symmetrix FCP Raid1
hdiskpower0     Available 06-08-02 PowerPath Device
hdiskpower1     Available 06-08-02 PowerPath Device
```

```
hdiskpower2     Available 06-08-02 PowerPath Device
hdiskpower3     Available 06-08-02 PowerPath Device
```

13. Type the **mkvdev** command for each **hdiskpower** device instance verifying the correct `vhostx` instances is being used as previously recorded.

    Input and output example:

```
$ mkvdev -vdev hdiskpower0 -vadapter vhost0
vtscsi0      Available

$ mkvdev -vdev hdiskpower1 -vadapter vhost1
vtscsi1      Available

$ mkvdev -vdev hdis

$ mkvdev -vdev hdiskpower3 -vadapter vhost1
vtscsi3      Available
```

14. Type the **lsmap** command to compare you previously recorded logical connections between the virtual target devices and the backing devices:

    ```
    $ lsmap -all
    ```

    Output example:

```
SVSA            Physloc                    Client PartitionID
--------------------------------------------------------------
vhost0          U9111.520.10DDEEC-V1-C20   0x00000002
VTD             vtscsi0
LUN             0x8100000000000000
Backing device  hdiskpower0
Physloc         U787B.001.DNW0F92-P1-C2-T1-L0

vhost1 U9111.520.10DDEEC-V1-C21            0x00000002
VTD             vtscsi1
LUN             0x8200000000000000
Backing device  hdiskpower1
Physloc         U787B.001.DNW0F92-P1-C2-T1-L1

VTD             vtscsi2
LUN             0x8300000000000000
Backing device  hdiskpower2
Physloc         U787B.001.DNW0F92-P1-C2-T1-L2

VTD             vtscsi3
LUN             0x8400000000000000
Backing device  hdiskpower3
Physloc         U787B.001.DNW0F92-P1-C2-T1-L3
```

15. Now that the Virtual I/O Client devices are exported, we can proceed in activating the Client Partition ID into the SMS menu and boot from the AIX operating system on virtual target device `vtscsi0` (`vtscsi0`, LUN 0x81, `hdiskpower0`).

Symmetrix PowerPath (PVID) to PowerPath (UDID) device migration

The following procedure outlines steps required for migrating devices from a PowerPath (PVID) implementation to a PowerPath (UDID) implementation. Considering there are different data formats associated between the PVID and UDID methods, it is highly recommended that all critical data be backed up before proceeding with these following steps.

It is also important to mention that you will need a total of free available space on the migration device that is equal too or greater than the used space on the source device due to the number of physical partitions being migrated.

This configuration consists of a Virtual I/O Server (VIOS) with Symmetrix, VNX series, and CLARiiON devices which are under the control of PowerPath 4.4 (PVID) and ODM Support Software version 5.2.0.1. Some but not all devices are also exported to the Virtual I/O client (VIOC) as virtual scsi devices.

1. It is important to gather all appropriate information before proceeding with the migration steps. Perform the following commands on the VIOS to record the configured devices.

   Log into the Virtual I/O Server partition and type the **`oem_setup_env`** command to access out of the restricted shell.

   Type the **`lsdev -Cc disk`** command to record your configured devices.

   > **`lsdev -Cc disk`**

   Output example:

```
hdisk0          Available 05-08-00-3,0      16 Bit LVD SCSI Disk Drive
hdisk1          Available 06-08-02 EMC Symmetrix FCP Raid1
hdisk2          Available 06-08-02 EMC Symmetrix FCP Raid1
hdisk3          Available 03-08-02 EMC Symmetrix FCP Raid1
hdisk4          Available 03-08-02 EMC Symmetrix FCP Raid1
hdisk5          Available 06-08-02 EMC CLARiiON FCP RAID 1/0 Disk
hdisk6          Available 06-08-02 EMC CLARiiON FCP RAID 1/0 Disk
hdisk7          Available 03-08-02 EMC CLARiiON FCP RAID 1/0 Disk
hdisk8          Available 03-08-02 EMC CLARiiON FCP RAID 1/0 Disk
hdiskpower0     Available 06-08-02 PowerPath Device
hdiskpower1     Available 06-08-02 PowerPath Device
hdiskpower2     Available 06-08-02 PowerPath Device
hdiskpower3     Available 06-08-02 PowerPath Device
```

2. Type the **`lspv`** command to display your configured devices.

   > **`lspv`**

   Output example:

```
hdisk0          00cfa11d15552dd5       rootvg       active
hdisk1          none                   None
hdisk2          none                   None
hdisk3          none                   None
hdisk4          none                   None
hdisk5          none                None
hdisk6          none                None
hdisk7          none                None
hdisk8          none                None
hdiskpower0     0cfa11d72d3aaa3     None
hdiskpower1     00cfa11d72d77343    None
hdiskpower2     00cfa11d72dde22c    None
hdiskpower3     00cfa11d72d53f81    None
hdiskpower4     00cfa11d72dba7e8    None
hdiskpower5     00cfa11d72dea14a    None
hdiskpower6     00cfa11d72dd8fdd    None
hdiskpower7     00cfa11d72d6eda6    None
```

3. Type the **`exit`** command to exit the shell.

   Then type the **`lsdev -virtual`** command to display your configured exported **`hdiskpower`** devices:

```
$ lsdev -virtual
```

Output example:

```
name     status              description
ent2     Available           Virtual I/O Ethernet Adapter (l-lan)
vhost0   Available           Virtual SCSI Server Adapter
vhost1   Available           Virtual SCSI Server Adapter
vtscsi0  Available           Virtual Target Device - Disk
vtscsi1  Available           Virtual Target Device - Disk
vtscsi2  Available           Virtual Target Device - Disk
vtscsi3  Available           Virtual Target Device - Disk
```

4. Type the **lsmap** command to record you logical connections between the virtual target devices and the backing devices:

```
$ lsmap -all
```

Output example:

```
SVSA               Physloc                    Client PartitionID
--------------------------------------------------
vhost0             U9111.520.10DDEEC-V1-C200x00000002
VTD                vtscsi0
LUN                0x8100000000000000
Backing device     hdiskpower0
Physloc            U787B.001.DNW0F92-P1-C2-T1-L0

vhost1             U9111.520.10DDEEC-V1-C21      0x00000002
VTD                vtscsi1
LUN                0x8200000000000000
Backing device     hdiskpower1
Physloc            U787B.001.DNW0F92-P1-C2-T1-L1

VTD                vtscsi2
LUN                   0x8300000000000000
Backing device        hdiskpower2
Physloc               U787B.001.DNW0F92-P1-C2-T1-L2

VTD                   vtscsi3
LUN                   0x8400000000000000
Backing device        hdiskpower3
Physloc               U787B.001.DNW0F92-P1-C2-T1-L3
```

5. Log in to the Virtual I/O Client and verify all volume groups are varied on. Record the PVID which resides on the volume groups by using the **lspv** command. You will notice that the PVID on the VIOC devices will be different than its backing device on the VIOS. This is because the existing configuration is using the physical volume identifier (PVID) method:

```
> lspv
```

Output example:

```
hdisk0       00cfa11d72d123a3 rootvg active
hdisk1       00cfa11d72d45643 vg1
hdisk2       00cfa11d72d7892c vg2
hdisk3       00cfa11d72dabc81 vg3
```

6. Type the **lsdev -Cc disk** command to display your configured devices. Once this is complete, shut down the client partition using the **shutdown** command.

Input and output example:

```
> lsdev -Cc disk
```

```
hdisk0      Available        Virtual SCSI Disk Drive
hdisk1      Available        Virtual SCSI Disk Drive
hdisk2      Available        Virtual SCSI Disk Drive
hdisk3      Available        Virtual SCSI Disk Drive
```

```
> shutdown
```

7. Log in to the Virtual I/O Server partition and type the **oem_setup_env** command to access out of the restricted shell. Download ODM support software version 5.2.0.3 or higher from the EMC FTP server and install the appropriate filesets.

   In this particular configuration we will be updating the Symmetrix AIX Support Software, Symmetrix Fibre Channel Support Software, VNX series Fibre Channel Software, and CLARiiON Fibre Channel Software.

8. Install PowerPath software version 4.4.2.2 or higher.

   Note: Version 4.4.2.2 requires a previous installation of version 4.4 and 4.4.2.

   Do not run the **powermt config** command after installing these versions.

   Once PowerPath is installed, reboot the VIOS partition.

9. Once the VIOS partition is in a running state, activate the previous VIOC partition.

10. Log in to the Virtual I/O Server and create a new virtual target device. This will be used as the target migration boot device.

    Type the **mkvdev** command for each boot device instance:

    ```
    $ mkvdev -vdev hdiskpower4 -vadapter vhost0
    ```

    Output example:

    ```
    vtscsi4 Available
    ```

11. Type the **mkvdev** command for each additional hdiskpower device instance verifying the correct vhostx instances is being used as previously recorded.

    Input and output example:

    ```
    $ mkvdev -vdev hdiskpower5 -vadapter vhost1
    vtscsi5     Available

    $ mkvdev -vdev hdiskpower6 -vadapter vhost1
    vtscsi6      Available

    $ mkvdev -vdev hdiskpower7 -vadapter vhost1
    vtscsi7      Available
    ```

12. Type the **lsdev** command to show the newly created Virtual Target Device:

    ```
    $ lsdev -virtual
    ```

    Output example:

```
name         status       description
ent2         Available    Virtual I/O Ethernet Adapter (l-lan)
vhost0       Available    Virtual SCSI Server Adapter
vhost1       Available    Virtual SCSI Server Adapter
vtscsi0      Available    Virtual Target Device - Disk
vtscsi1      Available    Virtual Target Device - Disk
vtscsi2      Available    Virtual Target Device - Disk
```

```
vtscsi3      Available    Virtual Target Device - Disk
vtscsi4      Available    Virtual Target Device - Disk
vtscsi5      Available    Virtual Target Device - Disk
vtscsi6      Available    Virtual Target Device - Disk
vtscsi7      Available    Virtual Target Device - Disk
```

13. Type the **lsmap** command to show the logical connections between the newly created devices:

    ```
    $ lsmap -all
    ```

    Output example:

```
SVSA            Physloc                    Client PartitionID
--------------------------------------------------------------
vhost0              9111.520.10DDEEC-V1-C20    0x00000002
VTD                 vtscsi0
LUN                 0x8100000000000000
Backing device      hdiskpower0
Physloc             U787B.001.DNW0F92-P1-C2-T1-L0

VTD                 vtscsi4
LUN                 0x8500000000000000
Backing device      hdiskpower4
Physloc             U787B.001.DNW0F92-P1-C2-T1-L4

vhost1              U9111.520.10DDEEC-V1-C210x00000002
VTD                 vtscsi1
LUN                 0x8200000000000000
Backing device      hdiskpower1
Physloc             U787B.001.DNW0F92-P1-C2-T1-L1

VTD                 vtscsi2
LUN                 0x8300000000000000
Backing device      hdiskpower2
Physloc             U787B.001.DNW0F92-P1-C2-T1-L2

VTD                 vtscsi3
LUN                 0x8400000000000000
Backing device      hdiskpower3
Physloc             U787B.001.DNW0F92-P1-C2-T1-L3

VTD                 vtscsi5
LUN                 0x8600000000000000
Backing device      hdiskpower5
Physloc             U787B.001.DNW0F92-P1-C2-T1-L5

VTD                 vtscsi6
LUN                 0x8700000000000000
Backing device      hdiskpower6
Physloc             U787B.001.DNW0F92-P1-C2-T1-L6

VTD             vtscsi7
LUN             0x8800000000000000
Backing device  hdiskpower7
PhyslocU        787B.001.DNW0F92-P1-C2-T1-L7
```

14. Now that the new Virtual I/O Client devices are exported, log in to the Virtual I/O Client and type the **cfgmgr** command to configure the new devices:

    ```
    > cfgmgr
    ```

15. Record the PVID which resides on the new devices by using the **lspv** command. You will notice that the PVID on the new VIOC device will be the same as its backing device on the VIOS. This is because the newly configured devices are using the unique device identifier (UDID) method.

    > **lspv**

    Output example:

```
hdisk0          00cfa11d72d123a3          rootvg       active
hdisk1          00cfa11d72d45643          vg1          active
hdisk2          00cfa11d72d7892c          vg2          active
hdisk3          00cfa11d72dabc81          vg3          active
hdisk4          00cfa11d72dba7e8          None
hdisk5          00cfa11d72dea14a          None
hdisk6          00cfa11d72dd8fdd          None
hdisk7          00cfa11d72d6eda6          None
```

16. Type the **lsdev -Cc disk** command to display your configured devices. Once this is complete, verify you volume groups are varied online:

    > **lsdev -Cc disk**

    Output example:

```
hdisk0          Available       Virtual SCSI Disk Drive
hdisk1          Available       Virtual SCSI Disk Drive
hdisk2          Available       Virtual SCSI Disk Drive
hdisk3          Available       Virtual SCSI Disk Drive
hdisk4          Available       Virtual SCSI Disk Drive
hdisk5          Available       Virtual SCSI Disk Drive
hdisk6          Available       Virtual SCSI Disk Drive
hdisk7          Available       Virtual SCSI Disk Drive
```

17. Type the **alt_disk_install or alt_disk_copy** command to create an alternate `rootvg` image to the new virtual device. This command can take a while to complete.

    Once complete, your boot device will be pointed to the `altinst_rootvg` copy. In this case the new boot device is hdisk4. Perform a system reboot once this process is complete.

    > **alt_disk_install -C hdisk4**
    or

    > **alt_disk_copy -d hdisk4**

    > **shutdown -Fr now**

18. Log in to the Virtual I/O Client and use the **extendvg** command to extend the existing volume groups to the newly configured devices:

    extendvg -f Vgname Pvname
    > **extendvg -f vg1 hdisk5**
    > **extendvg -f vg2 hdisk6**
    > **extendvg -f vg3 hdisk7**

19. Record the VG names and PVID that resides on the new devices by using the **lspv** command:

    > **lspv**

    Output example:

```
hdisk0      00cfa11d72d123a3       rootvg              active
hdisk1      00cfa11d72d45643       vg1                 active
```

```
hdisk2      00cfa11d72d7892c       vg2                active
hdisk3      00cfa11d72dabc81       vg3                active
hdisk4      00cfa11d72dba7e8       altinst_rootvg
hdisk5      00cfa11d72dea14a       vg1                active
hdisk6      00cfa11d72dd8fdd       vg2                active
hdisk7      00cfa11d72d6eda6       vg3                active
```

20. Type the **migratepv** command to move allocated physical partitions and the data to the extended volume group device:

```
migratepv SourcePhysicalVolume DestinationPhysicalVolume
> migratepv hdisk1 hdisk5
> migratepv hdisk2 hdisk6
> migratepv hdisk3 hdisk7
```

21. Type the **reducevg** command to remove original physical volume from the volume group:

```
reducevg VolumeGroup PhysicalVolume
> reducevg vg1 hdisk1
> reducevg vg2 hdisk2
> reducevg vg3 hdisk3
```

22. Record the VG names and PVID that resides on the new devices by using the **lspv** command:

    ```
    > lspv
    ```

    Output example:

```
hdisk0          00cfa11d72d123a3       rootvg             active
hdisk1          00cfa11d72d45643       vg1                active
hdisk2          00cfa11d72d7892c       vg2                active
hdisk3          00cfa11d72dabc81       vg3                active
hdisk4          00cfa11d72dba7e8       altinst_rootvg
hdisk5          00cfa11d72dea14a       vg1                active
hdisk6          00cfa11d72dd8fdd       vg2                active
hdisk7          00cfa11d72d6eda6       vg3                active
```

23. Vary off your volume groups using the **varyoffvg** command.

24. Remove the devices that were previously configured to the volume groups by using the **rmdev** command. Once this is complete, shut down the server using the **shutdown** command.

    ```
    > rmdev -dl hdisk1
    hdisk1 deleted
    > rmdev -dl hdisk2
    hdisk2 deleted
    > rmdev -dl hdisk3
    hdisk3 deleted
    > savebase
    > shutdown
    ```

25. Log in to the Virtual I/O Server and remove the PVID method configured exported devices using the **rmdev** command. This will unmap the logical connection between the virtual target devices and the backing devices:

    ```
    $ rmdev -dev vtscsi0
    vtscsi0 deleted
    $ rmdev -dev vtscsi1
    vtscsi1 deleted
    $ rmdev -dev vtscsi2
    vtscsi2 deleted
    ```

```
$ rmdev -dev vtscsi3
vtscsi3 deleted
```

26. Type the **lsdev** command to show the newly created Virtual Target Device:

```
$ lsdev -virtual
```

Output example:

```
name         status          description
ent2         Available       Virtual I/O Ethernet Adapter (l-lan)
vhost0       Available       Virtual SCSI Server Adapter
vhost1       Available       Virtual SCSI Server Adapter
vtscsi4      Available       Virtual Target Device - Disk
vtscsi5      Available       Virtual Target Device - Disk
vtscsi6      Available       Virtual Target Device - Disk
vtscsi7      Available       Virtual Target Device - Disk
```

27. Type the **lsmap** command to show the logical connections between the newly created devices:

```
$ lsmap -all
```

Output example:

```
SVSA            Physloc                         Client PartitionID
--------------------------------------------------------------
vhost0          9111.520.10DDEEC-V1-C20       0x00000002
VTD             tscsi4
LUN             0x8500000000000000
Backing device  diskpower4
Physloc         U787B.001.DNW0F92-P1-C2-T1-L4

vhost1          U9111.520.10DDEEC-V1-C21 0x00000002
VTD             vtscsi5
LUN             0x8600000000000000
Backing device  hdiskpower5
Physloc         U787B.001.DNW0F92-P1-C2-T1-L5


VTD             vtscsi6
LUN             0x8700000000000000
Backing device  hdiskpower6
Physloc         U787B.001.DNW0F92-P1-C2-T1-L6

VTD             vtscsi7
LUN             0x8800000000000000
Backing device  hdiskpower7
Physloc         U787B.001.DNW0F92-P1-C2-T1-L7
```

28. Now that the Virtual I/O Client devices are migrated, we can proceed in activating the Client PartitionID into the SMS menu and select the boot device. In this particular case the AIX operating system is on virtual target device `vtscsi4` (`vtscsi4`, LUN 0x85, `hdiskpower4`).

Shared Storage Pools with Virtual I/O Server

IBM PowerVM technology offers Shared Storage Pools. When a VIO client requires additional storage, the VIO server administrator can allocate storage from this pool. Shared Storage Pools use Cluster Aware AIX (CAA). CAA is an in-built clustering capability within AIX.

Shared Storage Pools offer another way of allocating storage to the VIO client through the VIO server, including NPIV, FILE, LUN, and LV provisioning.

In additional, the Shared Storage Pool feature comes with other features, including thin provisioning and snapshots, and can be configured to send out threshold alerts.

This section contains the following information:

Minimum requirements

The minimum requirements for a Shared Storage Pool is as follows:

- IBM PowerVM, Standard Edition
- VIOS version 2.2.0.11, Fix Pack 24, Service Pack 1 and later. The memory of VIO server should be no less than 4 GB.
- Minimum of two disks: One for the CAA repository and the other for the storage pool. Generally, the capacity of CAA repository disk should be at least 5 GB.
- Each VIO Server needs entries for others VIOS participating in the SSP Cluster in its /etc/hosts file. You must set up /etc/hosts file before creating Shared Storage Pool.

```
10.108.119.71aix119071aix119071.lss.emc.com
10.108.119.80aix119080aix119080.lss.emc.com
```

Note: There is a known issue when an `hdiskpower` device is configured as the repository disk and the underlying SAN is CLARiiON or Invista. Update VIO server to version 2.2.3.3 or later to avoid this issue.

Setting up Shared Storage Pool

To set up the Shared Storage Pool using the VIOS configuration menu, complete the following steps:

1. Present the LUNs from the SAN to both VIO servers.
2. Run the **chdev** command to set all the LUNs to no_reserve.

```
chdev -l hdiskpowerx -a reserve_policy=no_reserve
```

3. Run the **cfgassist** command to access the Config Assist for VIOS menu. Move the cursor to the **Shared Storage Pools** menu and press **Enter** to access the submenu.

```
Config Assist for VIOS

Move cursor to desired item and press Enter.
  Set Date and TimeZone
  Change Passwords
  Set System Security
  VIOS TCP/IP Configuration
  Install and Update Software
  Storage Management
  Devices
  Performance
  Role Based Access Control (RBAC)
  Shared Storage Pools              <--
  Electronic Service Agent
```

4. Move the cursor to **Manage Cluster and VIOS Modes** and press **Enter**.

```
  Manage Cluster and VIOS Nodes     <--
  Manage Storage Pools in Cluster
  Manage Logical Units in Storage Pool
```

5. Select **Create Cluster**.

```
Create Cluster                      <--
  List All Clusters
  Delete Cluster
  Add VIOS Node to Cluster
  Delete VIOS Node from Cluster
  List Nodes in Cluster
```

6. Enter the cluster name and storage pool name and select Repository disk and Storage Pool disks by pressing **Esc+4**. Then press **Enter** to create the cluster. This may take several minutes to complete.

```
Create Cluster

Type or select values in entry fields.
Press Enter AFTER making all desired changes.

[Entry Fields]
* Cluster name                                          [SSP4]
* Storage Pool name                                     [MyPool]
*PhysicalVolumesforRepository    [hdiskpower22]
+
  Physical Volumes for Storage Pool                     [hdiskpower8
hdiskpower9 hdiskpower10]
+
 FilenameinwhichPhysicalvolumesarespecified []
/

  Note:   *Cluster creation will take several
          minutes to complete.
          Please allow at least 10 minutes.
```

This can also be performed using the following commands:

```
cluster -create -clustername SSP4 -repopvs hdiskpower22 -spname
MyPool -sppvs hdiskpower8 hdiskpower9 hdiskpower10 -hostname
aix119080.lss.emc.com
```

Adding notes to the cluster

Return to the Manage Cluster and VIOS nodes menu and select **Add VIOS nodes to the cluster**. Enter the network name of Node to add. It may take several minutes to complete.

```
Add VIOS Node to Cluster

Type or select values in entry fields.
Press Enter AFTER making all desired changes.
[Entry Fields]
* Cluster name                                          SSP4
* Network name of Node to Add
   [aix119071.lss.emc.com]

  Note: * The network name may be either the
          hostname or IP address.
```

This can also be performed using the following commands:

```
cluster -addnode -clustername SSP4 -hostname aix119071.lss.emc.com
```

Checking the node status using CAA commands

The CAA commands, such as **lscluster**, can be used to check the node status of the cluster to ensure it is operational.

```
$ lscluster -i
Network/Storage Interface Query

Cluster Name: SSP4
Cluster UUID: 506d7f06-fc3a-11e3-8415-c54845457f80
Number of nodes reporting = 2
Number of nodes stale = 0
Number of nodes expected = 2

Node aix119080.lss.emc.com
Node UUID = 5084e3e4-fc3a-11e3-8415-c54845457f80
Number of interfaces discovered = 2
        Interface number 1, en3
                IFNET type = 6 (IFT_ETHER)
                NDD type = 7 (NDD_ISO88023)
                MAC address length = 6
                MAC address = 00:21:5E:EA:BC:E1
                Smoothed RTT across interface = 0
                Mean deviation in network RTT across interface = 0
                Probe interval for interface = 990 ms
                IFNET flags for interface = 0x1E084863
                NDD flags for interface = 0x0061081B
                Interface state = UP
                Number of regular addresses configured on interface = 1
IPv4 ADDRESS: 10.108.119.80 broadcast 10.108.119.255 netmask 255.255.255.0
Number of cluster multicast addresses configured on interface = 1
                IPv4 MULTICAST ADDRESS: 228.108.119.80
        Interface number 2, dpcom
                IFNET type = 0 (none)
                NDD type = 305 (NDD_PINGCOMM)
                Smoothed RTT across interface = 750
                Mean deviation in network RTT across interface = 1500
                Probe interval for interface = 22500 ms
                IFNET flags for interface = 0x00000000
                NDD flags for interface = 0x00000009
                Interface state = UP RESTRICTED AIX_CONTROLLED

Node aix119071.lss.emc.com
Node UUID = fef2c96e-fc3a-11e3-a843-c54845456541
Number of interfaces discovered = 2
        Interface number 1, en5
```

```
                IFNET type = 6 (IFT_ETHER)
                NDD type = 7 (NDD_ISO88023)
                MAC address length = 6
                MAC address = 00:21:5E:EA:BC:E0
                Smoothed RTT across interface = 0
                Mean deviation in network RTT across interface = 0
                Probe interval for interface = 990 ms
                IFNET flags for interface = 0x1E084863
                NDD flags for interface = 0x0061081B
                Interface state = UP
                Number of regular addresses configured on interface = 1
                IPv4 ADDRESS: 10.108.119.71 broadcast 10.108.119.255
                netmask   255.255.255.0
Number of cluster multicast addresses configured on interface = 1
                IPv4 MULTICAST ADDRESS: 228.108.119.80
        Interface number 2, dpcom
                IFNET type = 0 (none)
                NDD type = 305 (NDD_PINGCOMM)
                Smoothed RTT across interface = 750
                Mean deviation in network RTT across interface = 1500
                Probe interval for interface = 22500 ms
                IFNET flags for interface = 0x00000000
                NDD flags for interface = 0x00000009
                Interface state = UP RESTRICTED AIX_CONTROLLED
```

Checking the node and pool status

The following command can be used to check the node and pool status of the cluster.:

```
$ cluster -status -clustername SSP4
Cluster Name         State
SSP4                 OK

Node Name   MTM           Partition Num  State  Pool State
aix119080 9117-MMA021054845     1          OK      OK
aix119071 9117-MMA021054845     2          OK      OK
```

Creating and mapping logical unit to client

To create and map a logical unit to the client, complete the following steps:

1. Run the **cfgassist** command and move the cursor to the Shared Storage Pools menu and press **Enter** to access the submenu.
2. Move the cursor to **Manage Logical Unit in Storage Pool** and press **Enter**.

   ```
   Manage Cluster and VIOS Nodes
   Manage Storage Pools in Cluster
   Manage Logical Units in Storage Pool  <--
   ```

3. Select **Create and map Logical Unit**. Enter the necessary information to map the LUN to a specific VIO client.

   The Virtual Server Adapter information can be checked by running the **lsmap -all** command.

   By default SSP LUs are created with no_reserve attribute, so it is also ready for LPM.

   ```
   Create and Map Logical Unit

   Type or select values in entry fields.
   Press Enter AFTER making all desired changes.

                                                 [Entry Fields]
   * Cluster name                                      SSP4
   * Storage Pool name                                 MyPool
   * Logical Unit name                               [AIX119079_SSP]
   ```

```
* Logical Unit Size (MB)        [20480]
#
* Logical Unit Provision Type      [thin]
+#
* Virtual Server Adapter to Map     [vhost8]
+
  Virtual Target Device name                          [AIX119079_SSP]
```

When it is complete, the following information displays:

```
COMMAND STATUS

Command: OK            stdout: yes           stderr: no

Before command completion, additional instructions may appear below.

Lu Name:AIX119079_SSP
Lu Udid:9525822955013a8b4c3b51d3134e153a

Assigning logical unit 'AIX119079_SSP' as a backing device.

VTD:AIX119079_SSP
```

4. Make sure to map the logical unit on the second VIO server to ensure it has redundant paths.

Checking path status

The path status can be checked on VIO client. One path is from VIOS aix119071 and the other comes from VIOS aix119080.

```
# lspath|grep hdisk0

Enabled hdisk0  vscsi0
Enabled hdisk0  vscsi3
```

The following command can be used to list what is in the SSP:

```
$ lssp -clustername SSP4 -sp MyPool -bd

Lu Name      Size(mb) ProvisionType %Used Unused(mb)  Lu Udid
AIX119079_SSP 20480    THIN            0%   20481   9525822955013a8b4c3b51d3134e153a
```

Using snapshot and rollback

If an error occurs, you can easily snapshot and rollback the back devices with minimum outage.

1. Run the **cfgassist** command and move the cursor to the **Shared Storage Pools** menu and press **Enter** to access the submenu.
2. Move the cursor to **Manage Logical Unit in Storage Pool** and press **Enter**.

```
Shared Storage Pools

Move cursor to desired item and press Enter.

   Manage Cluster and VIOS Nodes
   Manage Storage Pools in Cluster
   Manage Logical Units in Storage Pool <--
```

3. Select **Create Logical Unit Snapshot**. You will be asked to select the logical units.

```
Select Logical Unit(s)
```

```
  Move cursor to desired item and press F7. Use arrow keys to scroll.
  ONE OR MORE items can be selected.
  Press Enter AFTER making all selections.

#LU NAME         LU UDID                                Size(MB)
AIX119079_SSP    9525822955013a8b4c3b51d3134e153a       20480

F1=Help    F2=Refresh    F3=Cancel
F1=Help    F7=Select     F8=Image     F10=Exit
F5=Reset   Enter=Do                   /=Find           n=Find Next
```

4. Select the logical unit and press **Enter** to complete the snapshot.

   You can also use the following command:

   snapshot -clustername SSP4 -create AIX119079_snapshot -spname MyPool -lu AIX119079_SSP

You can also rollback to the snapshot using either SMIT or CLI.

```
Manage Logical Units in Storage Pool

Move cursor to desired item and press Enter.

  Create and Map Logical Unit
  Create Logical Unit
  Map Logical Unit
  Unmap Logical Unit
  Delete Logical Unit
  List Logical Units
  List Logical Unit Maps
  Create Logical Unit Snapshot
  List Logical Unit Snapshots
  Rollback to Snapshot           ?
  Delete Snapshot

snapshot -clustername SSP4 -rollback AIX119079_snapshot -spname MyPool
   -lu AIX119079_SSP
```

New advanced features

The following new features were introduced in Shared Storage Pools, Phase 4:

- In previous Shared Storage Pools versions, physical volume can only be replaced with the following command:

```
$ pv -replace -clustername SSP_XIO -sp myPool -oldpv hdiskpower8 -newpv hdiskpower11
Current request action progress: % 5
Current request action progress: % 100
Given physical volume(s) have been replaced successfully.
```

Beginning in Shared Storage Pools, Phase 4, the physical volume can be removed from the pool.

```
$ pv -remove -clustername SSP_XIO -sp myPool -pv hdiskpower9
Given physical volume(s) have been removed successfully.
```

Although there is only one repository disk defined in Shared Storage Pools, this will not cause a single point of failure.

A damaged repository disk does not take down the Shared Storage Pools and it can be quickly rebuilt or replaced by running the following command:

```
$ chrepos -r -hdiskpower22,+hdiskpower14 -n SSP_XIO
chrepos: Successfully modified repository disk or disks.
```

You can even run this command when some nodes are down. The VIO servers will find the new repository disk and rejoin the cluster automatically.

- The **failgrp** command is also introduced in this new version. It is used to mirror the Shared Storage Pool across different SAN array. With this feature, even if the PVs in the pool are down, the Shared Storage Pools can still function normally.

  By default, the mirror status is NOT_MIRRORED.

```
$ failgrp -list
POOL_NAME: MyPool
TIER_NAME: SYSTEM
FG_NAME                              FG_SIZE(MB)  FG_STATE
Default                              68928        ONLINE

$ cluster -clustername SSP4 -status -verbose
Cluster Name:             SSP4
Cluster Id:               506d7f06fc3a11e38415c54845457f80
Cluster State:            OK
Repository Mode:          EVENT
Number of Nodes:          2
Nodes OK:                 2
Nodes DOWN:               0

    Pool Name:            MyPool
    Pool Id:              000000000A6C77500000000053AA7A34
Pool Mirror State:    NOT_MIRRORED
```

A mirror can be created with the following command:

```
failgrp -create -clustername SSP4 -sp MyPool -fg Backup:hdiskpower21 hdiskpower22 hdiskpower23
Backup FailureGroup has been created successfully.
```

To check the mirror status, use the following command:

```
$ cluster -status -clustername SSP4 -verbose
Cluster Name:             SSP4
Cluster Id:               506d7f06fc3a11e38415c54845457f80
Cluster State:            OK
Repository Mode:          EVENT
Number of Nodes:          2
Nodes OK:                 2
Nodes DOWN:               0
Pool Name:            MyPool
    Pool Id:              000000000A6C77500000000053AA7A34
    Pool Mirror State:    SYNCED

$ failgrp -list -verbose
POOL_NAME:MyPool
TIER_NAME:SYSTEM
FG_NAME:Default
FG_SIZE(MB):68928
FG_STATE:ONLINE

POOL_NAME:MyPool
TIER_NAME:SYSTEM
FG_NAME:Backup
FG_SIZE(MB):71248
FG_STATE:ONLINE
```

Logical partition/Virtual I/O Server (LPAR/VIOS) migration

This section assists AIX administrators migrate their existing AIX LPARs into a virtual environment while preserving existing data. After the introduction, the following information is included:

Introduction to LPAR/VIOS migration

Figure 5 on page 119 shows the initial setup used for the examples in this section. The LPAR is booting on AIX 5.3 and above, loaded with an ODM package, and uses multipathing software PowerPath 5.0 and above. The hdisk devices are color-coded to match the corresponding SAN disk. Both the boot disk and the data disk are SAN-attached. There is a specific reference to the Fibre Channel HBA card's physical location in the IO drawer. The AIX LPAR is booting from the Fibre Channel SAN.

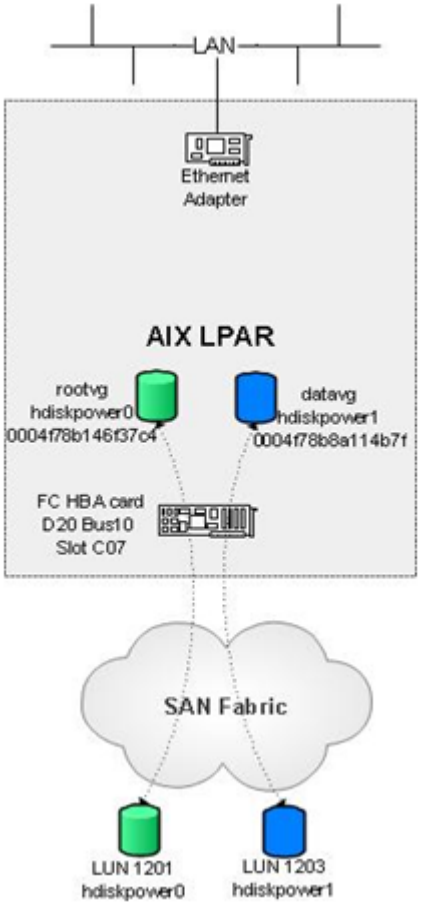


Figure 5 AIX LPAR initial setup example

The Virtual I/O Server will be set up as illustrated in Figure 6.

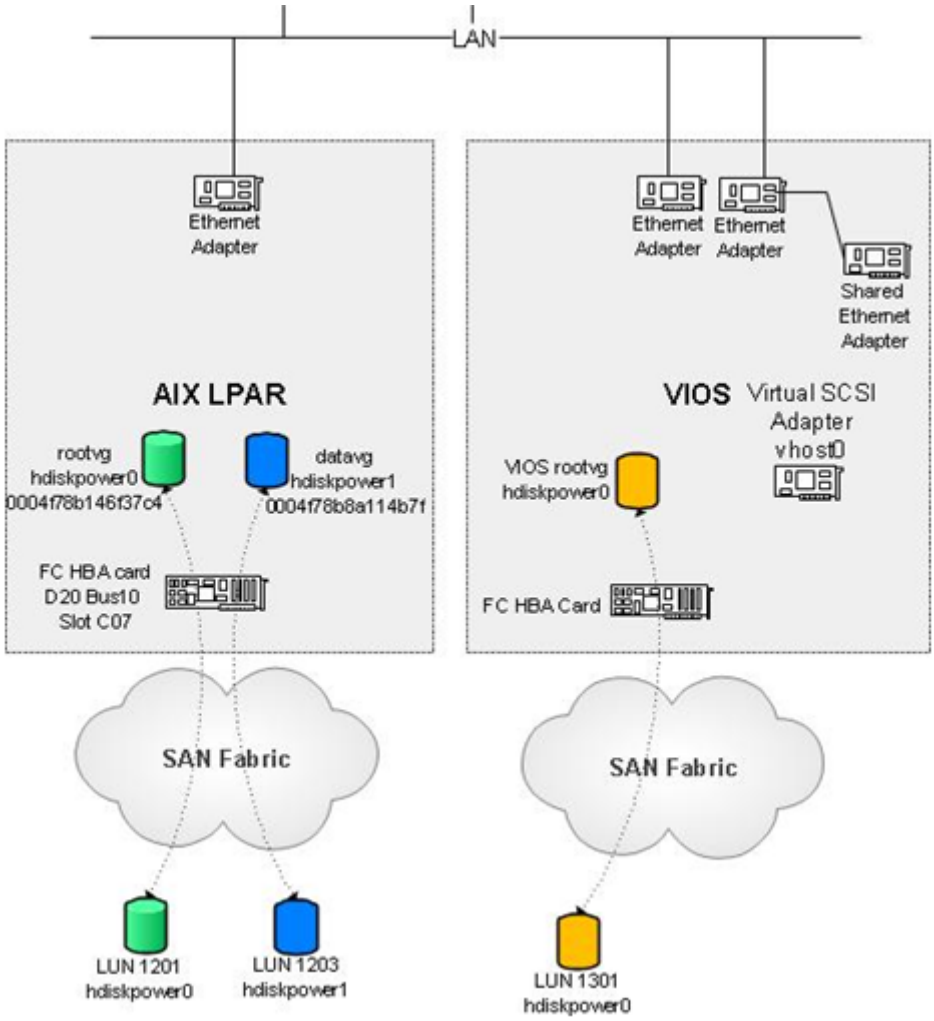


Figure 6 Virtual I/O Server setup example

The Virtual I/O Server is installed and booted from SAN. The AIX LPAR and the VIOS may, or may not, be on the same SAN fabric. In this section's examples, there is no differentiation and (unlike Figure 6) the reader can assume both of them are in the same fabric.

The final setup will look similar to that shown in Figure 7. This figure illustrates the relationship and interaction between all the LPARs involved in the migration. Notice that the hdisk devices are color-coded to refer to the same hard disk, although it is identified differently at the storage array and at different hosting partitions. The respective PVIDs (physical volume identifiers) are also recorded for verification purposes. Although the hdisk devices should be named in order after migrating, they may not be.

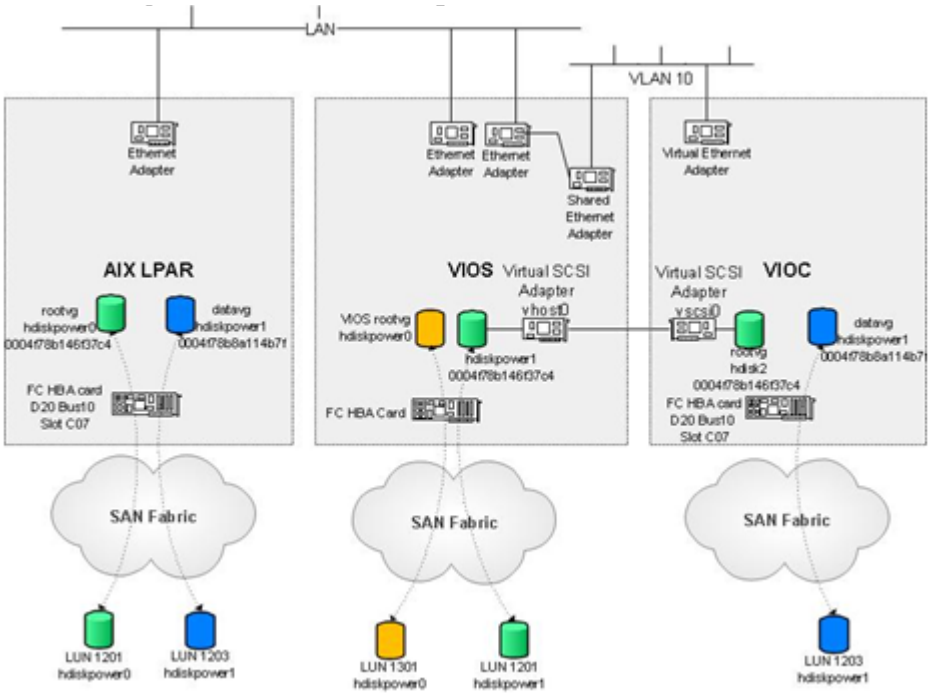


Figure 7 Final setup example

Prerequisites before migration

Before proceeding with the migration, some important information should be obtained from the existing system. It would be helpful to have documentation of the existing system setup. Refer to Table 5 on page 124 and Table 6 on page 125 for LPAR and hardware information for this migration example.

The following prerequisite steps will help to verify a successful migration.

Note: The LPAR must be booted on SAN for the remaining migration to take place.

From the LPAR:

1. Locate the PVID.

   Using the **lspv** command, find out from which hdisk the LPAR (hostname: SGELIBM22) is booting.

```
SGELIBM22/> lspv
hdisk0            none                        None
hdiskpower0       0004f78b8b2c79d9    rootvg          active
hdisk1            none                        None
hdiskpower1       0004f78b8a114b7f    datavg          active
```

   Notice that the LPAR is booting on **hdiskpower0**. Copy down the PVID of the `hdisk` devices.

2. Record the unique ID (UDID).

   There may be situations where the hdisk device sequence is different after the migration. Therefore, it is important to record the UDID information.

```
SGELIBM22/> odmget -q "name=hdiskpower0 and attribute=unique_id" CuAt | egrep
"name|attribute|value"
        name = "hdiskpower0"
        attribute = "unique_id"
          value = "3521360060160E4DF1E003454083E64D0DC1106RAID 503DGCfcp"
```

Repeat this step for all PowerPath devices.

```
SGELIBM22/> odmget -q "name=hdiskpower1 and attribute=unique_id" CuAt | egrep
"name|attribute|value"
        name = "hdiskpower1"
        attribute = "unique_id"
          value = "3521360060160E4DF1E0070933C0368D0DC1106RAID 503DGCfcp"
```

3. Discover the underlying storage array and volume information by using the **powermt display** command.

```
SGELIBM22/> powermt display dev=hdiskpower0
Pseudo name=hdiskpower0
CLARiiON ID=CK200073700702 [SGELIBM22]
Logical device ID=60060160E4DF1E003454083E64D0DC11 [LUN 1201]
state=alive; policy=CLAROpt; priority=0; queued-IOs=0
Owner: default=SP A, current=SP A       Array failover mode: 3
```

In this case, the array is booting from a CLARiiON system with ID CK200073700702, LUN 1201.

```
SGELIBM22/> powermt display dev=hdiskpower1
Pseudo name=hdiskpower1
CLARiiON ID=CK200073700702 [SGELIBM22]
Logical device ID=60060160E4DF1E0070933C0368D0DC11 [LUN 1203]
state=alive; policy=CLAROpt; priority=0; queued-IOs=0
Owner: default=SP A, current=SP A       Array failover mode: 3
```

The other `hdiskpower` device is LUN 1203.

4. Search for any existing mounted file systems.

These file systems may or may not reside in `rootvg`. After the migration, the host must still be able to read/write to these file systems. Therefore one way to do so is to run **df -k**.

```
SGELIBM22/> df -k

Filesystem    1024-blocks      Free %Used    Iused %Iused Mounted on

/dev/hd4          1114112   1092836    2%     1804     1% /
/dev/hd2          3473408   1332996   62%    33584    11% /usr
/dev/hd9var         65536     48072   27%      469     5% /var
/dev/hd3            65536     62144    6%       45     1% /tmp
/dev/hd1            65536     64380    2%      148     2% /home
/proc                   -         -    -        -     -  /proc
/dev/hd10opt        65536     13712   80%     1070    25% /opt
/dev/lv00         1048576   1015612    4%       18     1% /mnt/testfs1
/dev/fslv00       1048576   1048088    1%        6     1% /mnt/testfs2
```

In this example system, file system `testfs2` is created on `datavg` which resides on a separate disk. It is expected that the migrated host is able to read/write on all mounted file systems.

To prove that, some test files are created on the file systems:

```
SGELIBM22/> cd /mnt/testfs2
SGELIBM22/mnt/testfs2> vi this.file.is.created.on.fs2.before.migration
```

```
This file on FS2 is created BEFORE the migration
~
"this.file.is.created.on.fs2.before.migration" 1 line, 42 characters

SGELIBM22/mnt/testfs2> ls
lost+found
this.file.is.created.on.fs2.before.migration
```

5. Locate the World Wide Name and IP address.

   a. It is useful to know the World Wide Name (WWN) of the Fibre Channel HBA cards. This information is needed if there are SAN zoning changes. Two fcs ports are found using **`lsdev -Ccadapter`**. Use **lscfg** to extract the WWN.

```
SGELIBM22/> lscfg -vl fcs0 | grep Network
        Network Address.............10000000C96ABD02
SGELIBM22/> lscfg -vl fcs1 | grep Network
        Network Address.............10000000C96ABD03
```

   b. Extract the IP address information using **`ifconfig -a`** or **`smitty mktcpip`**.

```
* HOSTNAME                                           [SGELIBM22]
* Internet ADDRESS (dotted decimal)                  [10.32.133.22]
  Network MASK (dotted decimal)                      [255.255.254.0]
* Network INTERFACE                                   en0
  NAMESERVER
           Internet ADDRESS (dotted decimal)         [168.159.48.49]
           DOMAIN Name                               [lss.emc.com]
  Default Gateway
       Address (dotted decimal or symbolic name)     [10.32.132.1]
```

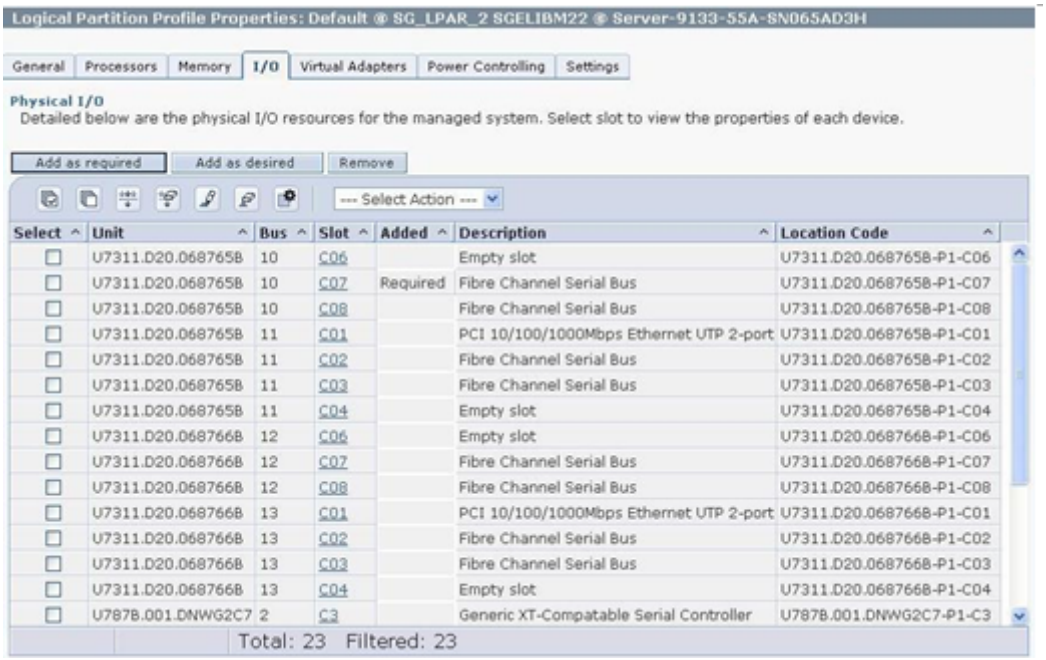
From the Hardware Managment Console (HMC):

1. Discover which physical components have been allocated to the existing LPAR by using the Hardware Management Console (HMC) to display the hardware profile the LPAR is running on.

2. Record all the settings and parameters in all the tabs under the profile properties.

Note: The administrator may want to adjust the processor and memory settings for the new Virtual I/O client partition.

The following figure shows that the LPAR uses a Fibre Channel HBA card on U7311.D20.068765B, bus 10, slot C07. Not shown in this figure is that the LPAR also has a Ethernet card; this information is not needed since the Ethernet connection would be virtualized



Information reference tables from LPAR for this migration example

Table 5 lists the LPAR information.

Table 5 LPAR information (1 of 2)

Device	Field	Value	Command	Check
hdiskpower0	PVID	0004f78b8b2c79d9*	lspv	
	UDID	3521360060160E4DF1E003454083E64D0DC11 06RAID 503DGCfcp*	odmget	
	Array	CK200073700702	powermt	
	LUN ID	LUN 1201		
hdiskpower1	PVID	0004f78b8a114b7f	lspv	
	UDID	3521360060160E4DF1E0070933C0368D0DC11 06RAID 503DGCfcp	odmget	
	Array	CK200073700702	powermt	
	LUN ID	LUN 1203		
fcs0	WWN	10000000C96ABD02	lscfg	
fcs1	WWN	10000000C96ABD03		

Table 5 LPAR information (2 of 2)

Device	Field	Value	Command	Check
IP	Hostname	SGELIBM22	smitty mktcpip	
	Address	10.32.133.22		
	Subnet	255.255.254.0		
	DNS	168.159.48.49		
	Gateway	10.32.132.1		
File Systems	FS name	testfs2	df –k lsvg –l [VGname]	
Volume Group	VG name	rootvg, datavg	lsvg	

Table 6 lists the hardware information.

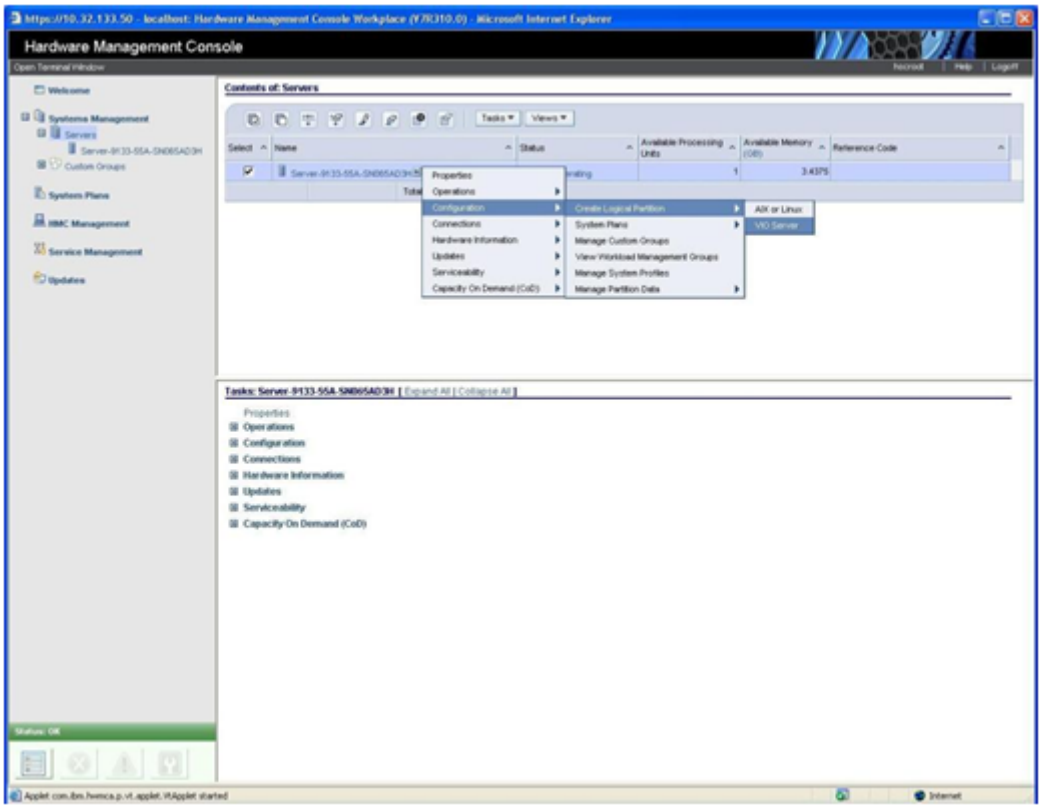
Table 6 Hardware information

Device	Unit	Bus	Slot
Fibre Channel HBA	U7311.D20.068765B	10	C07

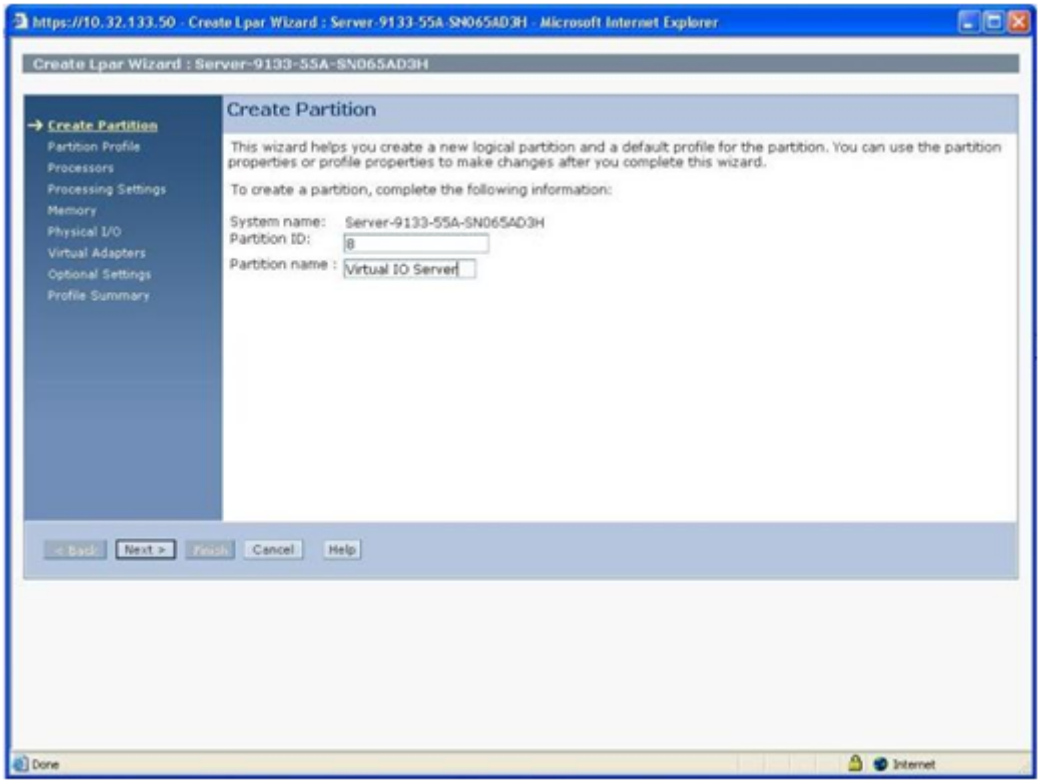
Setting up the Virtual I/O Server

A demonstration setup of a Virtual I/O Server (VIOS) is shown in this section.

1. Create a Virtual I/O server partition in the HMC.



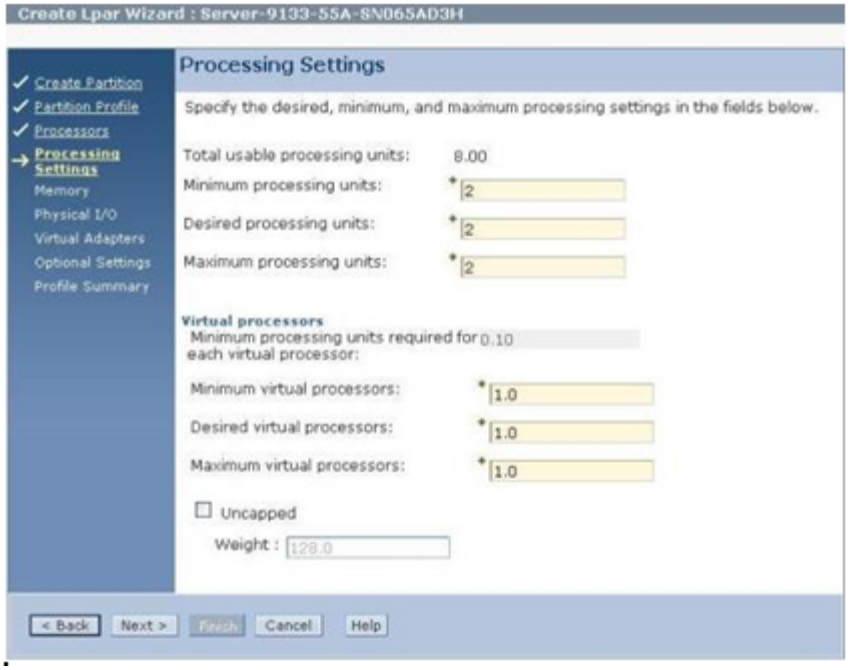
2. Name the partition. For this example, it is named Virtual I/O *Server*. The Partition ID is incremented automatically. Allow the HMC to assign the correct ID.



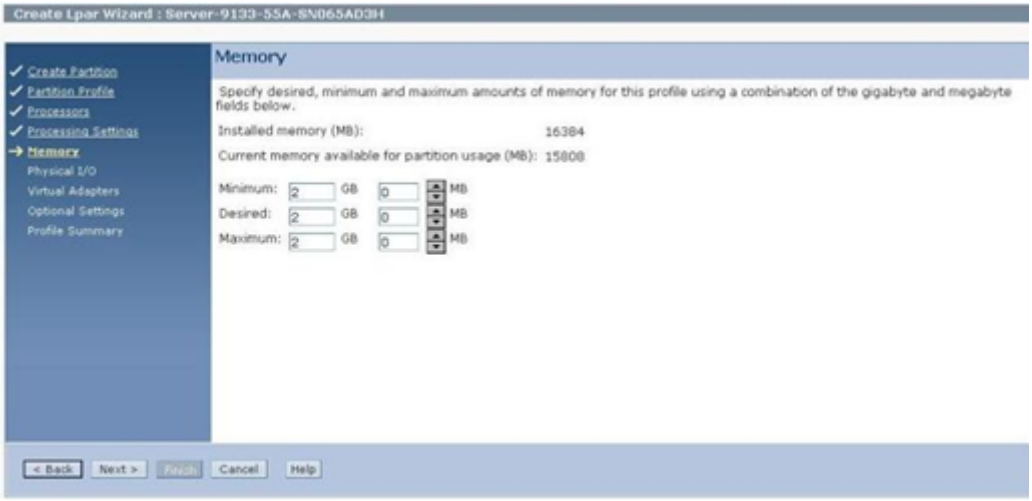
3. Name the partition profile **Default**.

4. Choose **Shared Processors**.

5. Set the **Processing Settings** as follows for demonstration sake. Adjust accordingly for different environment requirements.



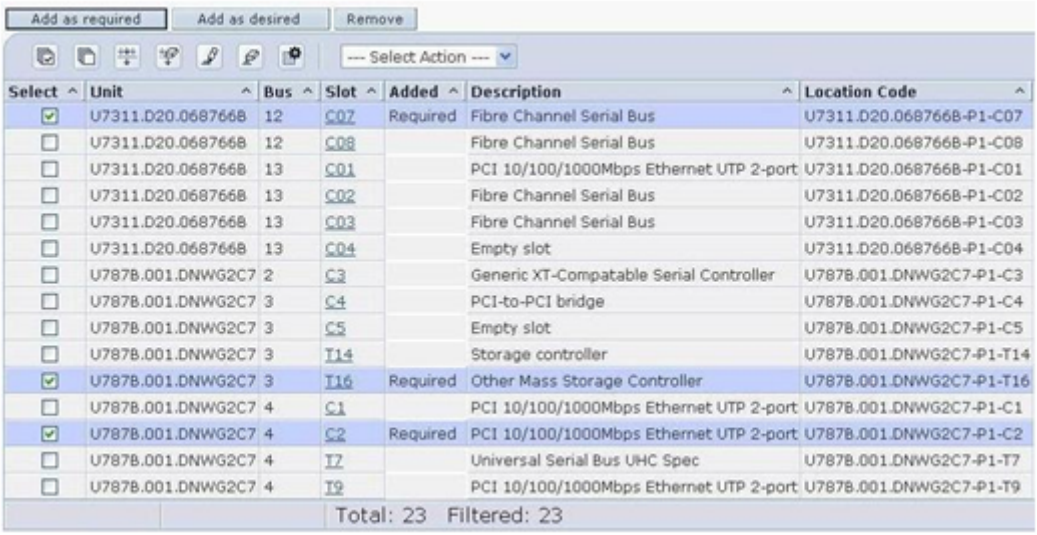
6. Assign an adequate amount of memory, adjusting accordingly to your needs.



7. Assign physical components to the VIOS.

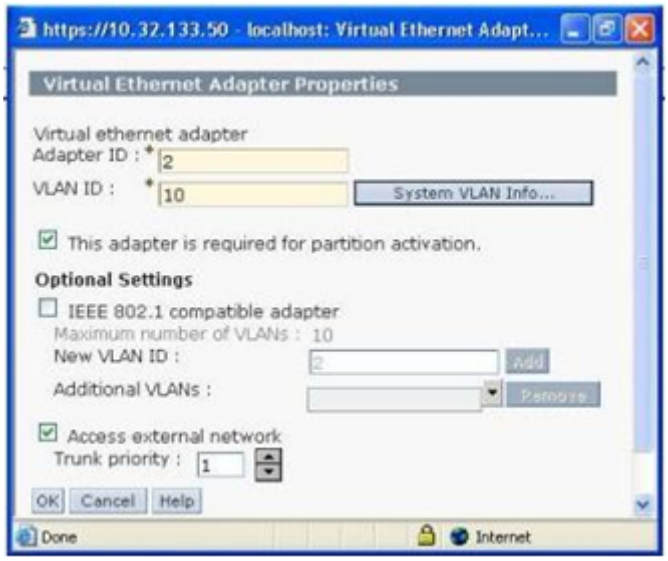
   a. Select the CD/DVD-ROM, a Fibre Channel (HBA) card, and an Ethernet card.

   b. After selecting the parts needed, click **Add as required** at the top of the window. Do *not* select components already assigned to other partitions.



8. Create a virtual Ethernet adapter in the next window.

a. Assign a VLAN ID (10 is used for this example). This information is needed for the Virtual I/O clients to connect to. Do *not* use any existing VLAN ID.



b. Create virtual SCSI adapters. These adapters are for Virtual I/O clients.

Note: It is a good practice to create more adapters than required. This way, the server does not need to shut down to create additional adapters to accommodate more clients. Since the information about the client is not yet available, leave **Any client partition can connect** checked. If there is client information available, assign and fix the connecting client.

The following figure shows a summary of the virtual adapters. An additional virtual SCSI adapter has been created as spare.

9. Complete the profile configuration.

A profile summary displays.

Installing the Virtual I/O Server

The Virtual I/O Server requires a Fibre Channel (HBA) card assigned to it in order to attach the SAN disk for boot support. Complete the following steps:

1. Zone the HBA card to see the storage port.
2. Using tools such as Symmask or Unisphere/Navisphere Manager (depending on storage array), create a 40 gigabyte SAN-attached disk for the Virtual I/O Server.
3. Once the Virtual I/O Server partition is complete, activate the partition and enter the System Management Services (SMS) mode.
4. Insert the Virtual I/O Server installation DVD into the DVD-ROM drive and select the boot option to use the CD/DVD-ROM.
5. Restart the Virtual I/O Server by exiting the SMS mode and selecting to reboot in normal mode.

   When the Virtual I/O Server is rebooting from the DVD, it should be able to install unto the SAN disk that was created.
6. Once the installation is complete, set the IP address of the Virtual I/O Server used to administer it.

   The Virtual I/O Server operating system runs on a restricted access mode by default. To use a non-restricted access mode, use the **oem_setup_env** command to access commands available to root users in AIX.

   ```
   $ oem_setup_env
   # smitty mktcpip
   ```
7. From the **smitty** menu, set the IP address for the Virtual I/O Server. This IP address will be used to access the Virtual I/O Server.
8. Install PowerPath software on the Virtual I/O Server. For more details on installing PowerPath on your Virtual I/O Server, download the *PowerPath for AIX Installation and Administration Guide* at Dell EMC Online Support.

Setting up the Virtual I/O Client

This section relies heavily on the LPAR information gathered in the steps under "Prerequisites before migration" on page 121. The Virtual I/O Client (VIOC) will need to have the same physical make up of the existing LPAR. Refer to Table 6 on page 125 and Table 7 on page 136 for LPAR and hardware information for this migration example.

To create the Virtual I/O Client partition:

1. Shut down the existing LPAR and remove the physical components from the LPAR hardware profile that will be required to activate the Virtual I/O Client.

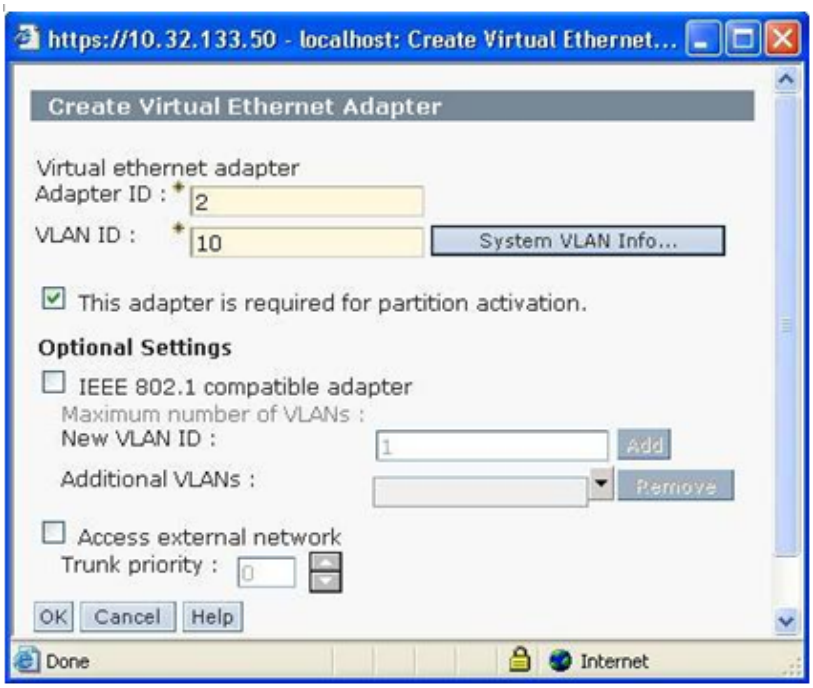
2. Follow the steps listed in "Setting up the Virtual I/O Server" on page 125.

Note: Adjust the processor and memory requirements accordingly in the respective sections. For this example, the '**`VIOC-Migrate-Test`**' client partition was created.

---

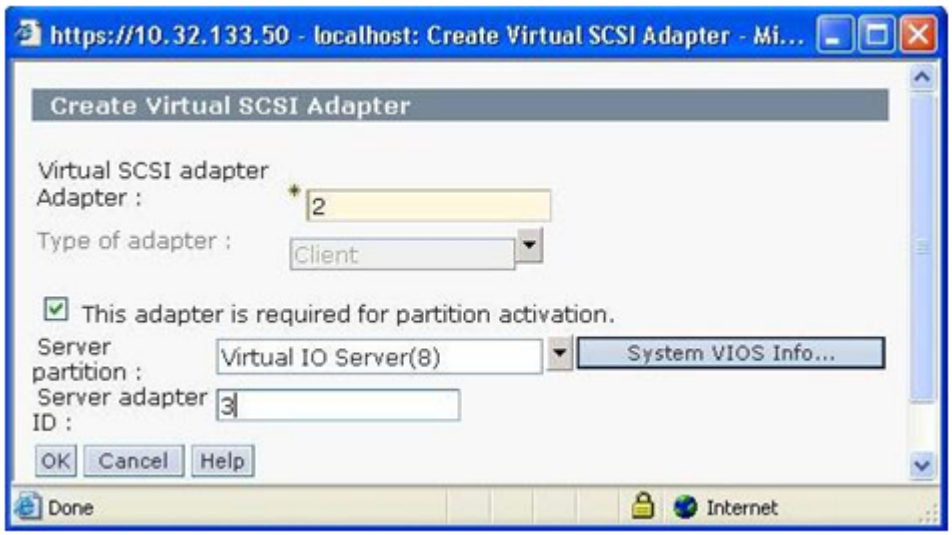
3. Unless the LPAR has a existing Fibre Channel HBA card that has to be added back to the Virtual I/O Client, skip the Physical I/O section and do not add any physical components to the Virtual I/O Client. In this case, the HBA card is still required as a non-boot disk (`hdiskpower1`) is on the SAN.
4. In the Virtual I/O section, create a Virtual Ethernet adapter.

Note: The VLAN ID is the same as the VLAN ID that was assigned for the Virtual I/O Server. Do *not* check the **Access external network** option for this Virtual I/O Client.



5. Create a Virtual SCSI adapter for the Virtual I/O Client.

   Be precise with the Server partition and the Server adapter ID that are going to connect the client adapters. Always check the Server's profile to find any unused adapter IDs and then connect the correct Server adapter ID. In this example, the client's Virtual SCSI adapter 2 will always connect with our Virtual I/O Server's adapter 3, as shown in the next figure.



6. **Optional:** Create an additional Virtual Ethernet adapter and a Virtual SCSI adapter to provision for the dual VIOS setup.

Configuring the Virtual I/O Server (VIOS) and Virtual I/O client (VIOC)

To set up the virtual SCSI:

The Virtual I/O Client accesses its SAN boot disk through the Virtual I/O Server. Therefore, the Virtual I/O Server needs to be able to access the LPAR's SAN boot disk. To allow the Virtual I/O Server access, follow these steps:

1. Shutdown the LPAR to prevent any possible data corruption. Since two hosts have concurrent access, data can be corrupted.
2. Use **symmask** or add the LPAR's SAN boot disk to the Virtual I/O Server's Storage Group, removing it from its original group.
3. Run #**cfgmgr** on the Virtual I/O Server and check that the boot disk has been added to the hard disk device list by using #**lsdev -Ccdisk**.

   Run the **$oem_setup_env** command if the prompt is still in the restricted access mode of the Virtual I/O Server.

   a. Before running #**cfgmgr**, verify that hdisk0 is the disk that the Virtual I/O Server is installed on.

```
# lsdev -Ccdisk
hdisk0      Available 05-08-02 EMC CLARiiON FCP RAID 5 Disk
hdiskpower0 Available 05-08-02 PowerPath Device
```

   b. After running #**cfgmgr**, verify that `hdisk1` and `hdiskpower1` are added to the hard disk device list.

      In this example, we only have a single zoned path to the storage port; therefore there is only one additional hdisk device. If there are multiple paths, multiple hdisk devices will display, but only one `hdiskpower` device will be created.

```
# cfgmgr
# lsdev -Ccdisk
hdisk0        Available 05-08-02 EMC CLARiiON FCP RAID 5 Disk
hdisk1        Available 05-08-02 EMC CLARiiON FCP RAID 5 Disk
hdiskpower0 Available 05-08-02 PowerPath Device
hdiskpower1 Available 05-08-02 PowerPath Device
```

4. Enter **#exit** and use the restricted access mode. Run **$lsdev -virtual** and the following displays:

```
$ lsdev -virtual
name            status     description
ent2            Available  Virtual I/O Ethernet Adapter (l-lan)
vhost0          Available  Virtual SCSI Server Adapter
vhost1          Available  Virtual SCSI Server Adapter
vsa0            Available  LPAR Virtual Serial Adapter
```

   The vhost0 and vhost1 are the Virtual SCSI adapters created using the HMC.

5. Connect the LPAR boot disk to the first Virtual SCSI adapter by returning to the restricted access mode and issuing the following command:

```
$ mkvdev -vdev hdiskpower1 -vadapter vhost0 -dev vioc_mig_lun
vioc_mig_lun Available
```

The `hdiskpower1` device is now linked to the virtual SCSI adapter vhost0, and a Virtual Target Device (VTD) named `vioc_mig_lun` is created. This VTD is available to the connecting Virtual I/O Client. The prompt indicates a successful creation of the Virtual Target Device by showing the device as **Available**.

6. Verify the successful creation of the Virtual Target Device by running:

   ```
   $ lsmap -vadapter vhost0
   ```

   vhost0 is now linked to `hdiskpower1` and the Virtual Target Device name is `vioc_mig_lun`.

```
SVSA            Physloc                     Client Partition ID
--------------- --------------------------  ------------------
vhost0          U9133.55A.065AD3H-V8-C3     0x00000000
VTD               vioc_mig_lun
LUN               0x8100000000000000
Backing device    hdiskpower1
Physloc           U7311.D20.068766B-P1-C07-T1-L2
```

To set up the Virtual Ethernet:

Create a Shared Ethernet Adapter (SEA, required for the Virtual I/O Client to connect to, by completing the following steps:

1. Run **`$ lsdev -virtual`** to verify that the Ethernet trunk adapter is available.

```
$ lsdev -virtual
name             status     description
ent2             Available  Virtual I/O Ethernet Adapter (l-lan)
vhost0           Available  Virtual SCSI Server Adapter
vhost1           Available  Virtual SCSI Server Adapter
vsa0             Available  LPAR Virtual Serial Adapter
vioc_mig_lun  Available  Virtual Target Device - Disk
```

2. Use **`$ lsdev -type adapter`** to look for another available ent adapter.

   Note: Previously, **`ent0`** was used for the Virtual I/O Server's IP address. Do *not* reuse it to create the SEA. In this example, use **`ent1`**.

```
$ lsdev -type adapter
name             status     description
ent0             Available  2-Port 10/100/1000 Base-TX PCI-X Adapter
ent1             Available  2-Port 10/100/1000 Base-TX PCI-X Adapter
ent2             Available  Virtual I/O Ethernet Adapter (l-lan)
fcs0             Available  FC Adapter
fcs1             Available  FC Adapter
ide0             Available  ATA/IDE Controller Device
vhost0           Available  Virtual SCSI Server Adapter
vhost1           Available  Virtual SCSI Server Adapter
vsa0             Available  LPAR Virtual Serial Adapter
```

3. Issue the following command to create the SEA:

```
$ mkvdev -sea ent1 -vadapter ent2 -default ent2 -defaultid 2
```

where:

- The default field in this setup is ent2
- The default id is the adapter ID that the HMC gave to the Virtual Ethernet port created.

A successful run will return a newly created Shared Ethernet Adapter, **ent3**:

```
ent3 Available
en3
et3
```

4. Check that the Shared Ethernet Adapter has been created on the Virtual I/O Server.

```
$ lsdev -virtual
name             status      description
ent2             Available   Virtual I/O Ethernet Adapter (l-lan)
vhost0           Available   Virtual SCSI Server Adapter
vhost1           Available   Virtual SCSI Server Adapter
vsa0             Available   LPAR Virtual Serial Adapter
vioc_mig_lun  Available  Virtual Target Device - Disk
ent3             Available   Shared Ethernet Adapter
```

5. Run **$ lsmap** to verify:

```
$ lsmap -all -net
SVEA     Physloc
------   --------------------------------------
ent2     U9133.55A.065AD3H-V8-C2-T1
SEA               ent3
Backing device    ent1
Physloc           U787B.001.DNWG2C7-P1-C2-T2
```

6. Use the non-restricted access mode and set the IP for the new adapter, **ent3**. The bridge to connect the Virtual I/O Clients from the VLAN to the enterprise is now configured.

```
$ oem_setup_env
# smitty mktcpip
```

The IP of **ent3** should be different from the IP address of the Virtual I/O Server.

Information reference tables for Virtual I/O client and server for this example

Table 7 lists the VIOC information.

Table 7 VIOC information (1 of 2)

Device	Field	Value	Command	Check
hdisk2	PVID	0004f78b8b2c79d9*	lspv	
	UDID	3521360060160E4DF1E003454083E64D0DC 1106RAID 503DGCfcp*	odmget	
	Array	-N/A-	powermt	
	LUN ID	-N/A-		

Table 7 VIOC information (2 of 2)

Device	Field	Value	Command	Check
hdiskpower1	PVID	0004f78b8a114b7f	lspv	
	UDID	3521360060160E4DF1E0070933C0368D0DC 1106RAID 503DGCfcp	odmget	
	Array	CK200073700702	powermt	
	LUN ID	LUN 1203		
fcs0	WWN	10000000C96ABD02	lscfg	
fcs1	WWN	10000000C96ABD03		
IP	Hostname	SGELIBM22	smitty mktcpip	
	Address	10.32.133.22		
	Subnet	255.255.254.0		
	DNS	168.159.48.49		
	Gateway	10.32.132.1		
File Systems	VG	testfs2	df –k lsvg –l [VGname]	
Volume Group	VG name	rootvg, datavg	lsvg	

Table 8 lists the hardware information.

Table 8 Hardware information

Device	Unit	Bus	Slot
Fibre Channel HBA	U7311.D20.068765B	10	C07

Table 9 lists the VIOS information.

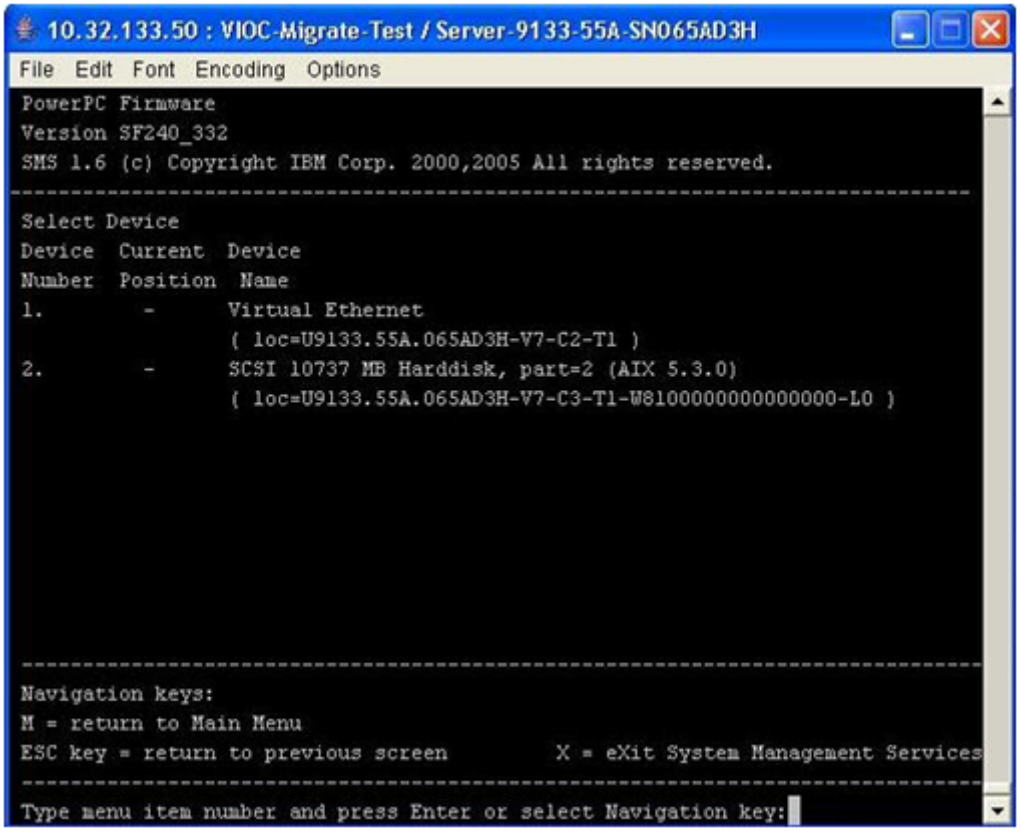
Table 9 VIOS information

Device	Field	Value	Command	Check
vhost0	Backing Device	hdiskpower1	lsmap -all	
	VTD	vioc_mig_lun		
hdiskpower1	PVID	0004f78b8b2c79d9*	lspv	
	UDID	3521360060160E4DF1E003454083E64D0D C1106RAID 503DGCfcp*	odmget	
	Array	CK200073700702	powermt	
	LUN ID	LUN 1201		

Starting the Virtual I/O Client

To start the Virtual I/O Client, follow these steps:

1. Activate the Virtual I/O Client from the HMC.
2. Enter the System Management Services (SMS) mode on startup to scan for bootable hard disk devices.
3. From the boot options, find the SCSI disk that has your boot image, select that device number, and boot from it.



Once the client has been activated, users will not be able to telnet into your server because the IP address of the new Virtual Ethernet adapter on the Virtual I/O Client has not been configured yet.

4. Set the IP for the Virtual I/O Client.
   a. Enter **`lsdev -Ccadapter.`** The host has a new Virtual I/O Ethernet Adapter. Previous Ethernet adapters are now marked as **Defined**.

```
SGELIBM22/> lsdev -Ccadapter
ent0   Available       Virtual I/O Ethernet Adapter (l-lan)
ent3   Defined   05-08 2-Port 10/100/1000 Base-TX PCI-X Adapter
ent4   Defined   05-09 2-Port 10/100/1000 Base-TX PCI-X Adapter
fcs0   Available 06-08 FC Adapter
fcs1   Available 06-09 FC Adapter
ide0   Defined   04-08 ATA/IDE Controller Device
vsa0   Available       LPAR Virtual Serial Adapter
vscsi0 Available       Virtual SCSI Client Adapter
```

   b. Set the IP of this new **`ent0`** Virtual Ethernet Adapter by using **`smitty mktcpip`** to configure it with the same settings as the LPAR.
   c. Ping the gateway to verify it can get a response.

5. Verify the hard disk devices.

   The old hard disk, hdisk0, is now marked as **Defined**. A new Virtual SCSI disk, hdisk2, now exists.

```
SGELIBM22/> lsdev -Ccdisk
hdisk0      Defined   06-08-02 EMC CLARiiON FCP RAID 5 Disk
hdisk1      Available 06-08-02 EMC CLARiiON FCP RAID 5 Disk
hdisk2      Available          Virtual SCSI Disk Drive
hdisk3      Available 06-08-02 EMC CLARiiON FCP LUNZ Disk
hdiskpower0 Defined   06-08-02 PowerPath Device
hdiskpower1 Available 06-08-02 PowerPath Device
```

6. Enter **`lspv`**.

```
SGELIBM22/> lspv
hdisk1          none                None
hdiskpower1     0004f78b8a114b7f    datavg          active
hdisk2          0004f78b8b2c79d9    rootvg          active
hdisk3          none                None
```

   Note that the `rootvg` is no longer on `hdiskpower0` but is now on `hdisk2`. Also note that the PVIDs of the hdisks are not changed.

7. Manually change the operating system's bootlist. Currently, it still points to the old `hdiskpower0` or `hdisk0` where the `rootvg` was located. Changing it avoids wasted time trying to find the `rootvg` on a non-existing hdisk when the host reboots.

```
SGELIBM22/> bootlist -m normal -o
hdisk0
SGELIBM22/> bootlist -m normal hdisk2
SGELIBM22/> bootlist -m normal -o
hdisk2 blv=hd5
```

8. Issue the **`lspath`** command to discover to which Virtual SCSI adapter on the host the hdisk2 is connected.

```
SGELIBM22> lspath
Enabled hdisk2 vscsi0
```

Finalizing the migration

Perform a final verification by completely the following steps.

1. Ensure that all the file systems that were previously mounted are correctly mounted and users can read/write on them.

```
SGELIBM22/> cd /mnt/testfs2
SGELIBM22/mnt/testfs2> ls
lost+found
this.file.is.created.on.fs2.before.migration
SGELIBM22/mnt/testfs2> pg this.file.is.created.on.fs2.before.migration
This file on FS2 is created BEFORE the migration
(EOF):

SGELIBM22/mnt/testfs2>
```

2. Write to the file systems.

```
SGELIBM22/mnt/testfs2> ls
lost+found
this.file.is.created.on.fs2.before.migration
```

```
SGELIBM22/mnt/testfs2> vi this.file.is.created.on.fs2.before.migration
"this.file.is.created.on.fs2.before.migration" 1 line, 42 characters
This file is created BEFORE the migration
This line is added AFTER the migration
~
"this.file.is.created.on.fs2.before.migration" 2 lines, 81 characters

SGELIBM22/mnt/testfs2>
```

Reading and writing on the file systems that existed before the migration is successful.

3. Check that the PVIDs of the `hdisk` and `hdiskpower` devices remained the same by comparing the PVIDs (copied as instructed in Step 1 on page 121), with the output from **lspv**.
4. Check that the UDID remained the same (copied as instructed in Step 2 on page 121), and compare it to the Virtual I/O Server's backing device.
   - On the Virtual I/O Client:

```
SGELIBM22/> odmget -q "name=hdisk2 and attribute=unique_id" CuAt | egrep "name|attribute|value"
        name = "hdisk2"
        attribute = "unique_id"
        value = "4A3533521360060160E4DF1E003454083E64D0DC1106RAID    503DGCfcp05VDASD03AIXvscsi"
```

   - On the Virtual I/O Server:

```
# odmget -q "name=hdiskpower1 and attribute=unique_id" CuAt | egrep "name|attribute|value"
        name = "hdiskpower1"
        attribute = "unique_id"
        value = "3521360060160E4DF1E003454083E64D0DC1106RAID 503DGCfcp"
```

Note: After checking that they have the same UDID, also check the records that were made before the migration. These should be the same.

Best practice

When starting the Virtual I/O Client, the physical devices that were previously owned by the server are now unavailable. It is recommended that you remove all these devices from the list to avoid confusion. Use **rmdev -Rdl** on these devices to remove them from the host devices list.

Implementing a dual Virtual I/O Server

The purpose of the implementing a dual Virtual I/O Server (VIOS) is for redundancy. In the single Virtual I/O Server setup, there are single points of failure that could cause an outage on the Virtual I/O Client. The failure of the Virtual I/O Server would render all the Virtual I/O Clients that are sharing resources from the Virtual I/O Server to malfunction. To avoid this, it is recommended that you set up a dual Virtual I/O Server system configuration.

Figure 8 on page 141 illustrates the desired setup for dual VIOS implementation. This design will withstand the failure of a single Virtual I/O Server. Setting up the second Virtual I/O Server is exactly the same as described in "Setting up the Virtual I/O Server" on page 125.

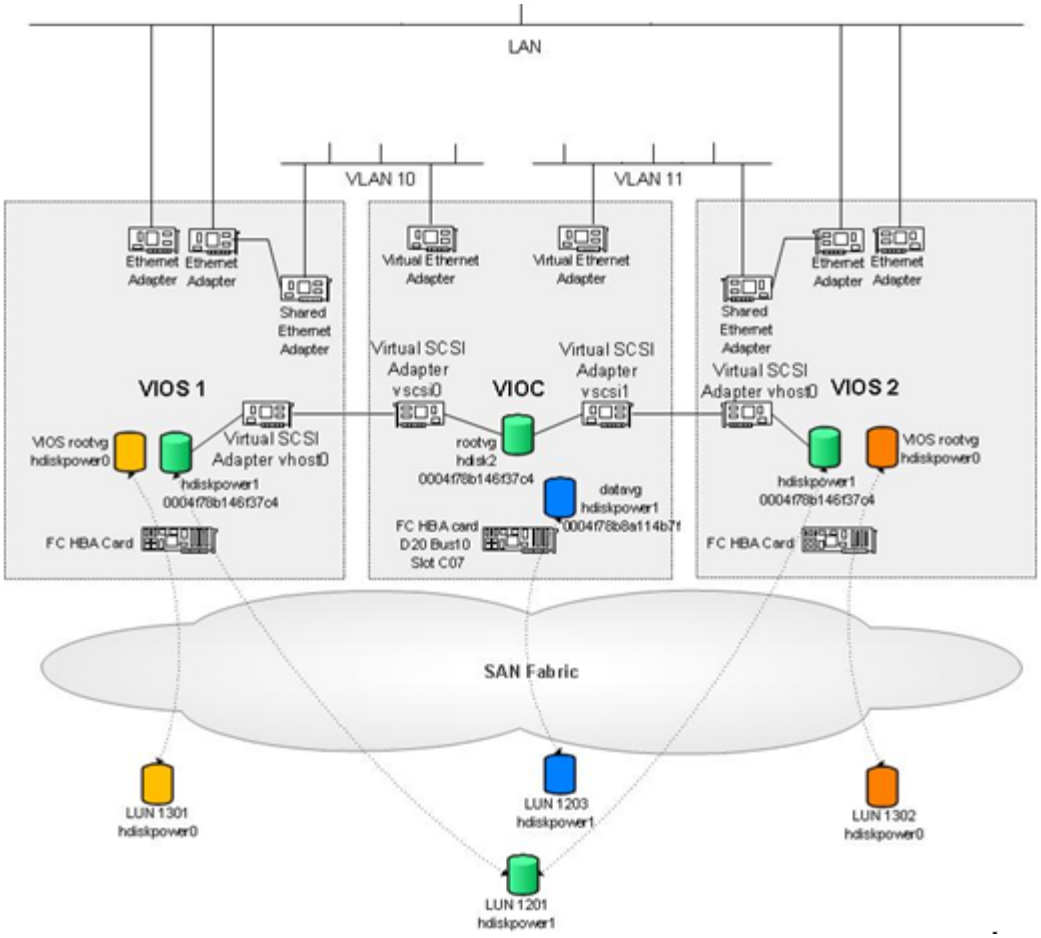


Figure 8 Dual VIOS example

Note the VLAN ID (in this example, ID 11 is assigned) of the new Virtual I/O Server and the Virtual SCSI adapter ID that is intended to pair.

Before proceeding with the setup, enter **`#lspath`** on the Virtual I/O Client and verify that `rootvg` on `hdisk2` can only be accessed via one route.

```
SGELIBM22/> lspath
Enabled hdisk2 vscsi0
```

On the VIOS:

1. To configure the SAN connection, configure the FC SCSI attributes on both of the Virtual I/O Servers as follows:

```
$ chdev -dev fscsi0 -attr fc_err_recov=fast_fail dyntrk=yes -perm
$ chdev -dev fscsi1 -attr fc_err_recov=fast_fail dyntrk=yes -perm
```

IMPORTANT

Configure these settings for high availability and quick failover. The Virtual I/O Server needs to be rebooted for the changes to take effect.

2. The hard disk will be made accessible to both Virtual I/O Servers. Configure the hard disk to allow access from multiple Virtual I/O servers.
3. After zoning and presenting the LUN 1201 (that the Virtual I/O Client boots on) to the second Virtual I/O Server, check its attributes.

```
$ lsdev -dev hdiskpower1 -attr
attribute     value              description   user_settable
...
reserve_lock  yes            Reserve device on open     True
...
```

4. Change the `reserve_lock` attribute to **no**.

```
$ chdev -dev hdiskpower1 -attr reserve_lock=no
```

In PowerPath v5.5 and later, reserve_policy = no_reserve.

```
$ chdev -dev hdiskpower1 -attr reserve_policy=no_reserve
```

5. Change the attribute on both of the Servers.
6. After this attribute is changed, add the `hdiskpower` device to a Virtual Adapter on the second Virtual I/O Server.

```
$ mkvdev -vdev hdiskpower1 -vadapter vhost0 -dev vioc_mig_lun
vioc_mig_lun Available
```

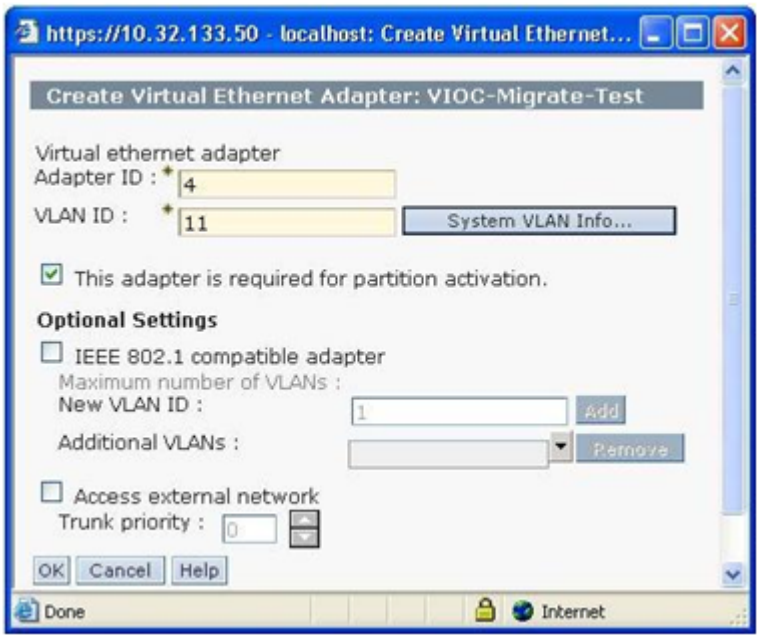
7. If there are no extra Virtual SCSI adapters free for connection on the second Virtual I/O Server, add them into its hardware profile through the HMC.
8. Configure the Virtual SCSI adapter to allow *any* client to connect, or manually configure it to allow only **VIOC-Migrate-Test** to connect. Currently there is no information of the adapter ID of the Client. It can be edited after creating the additional adapter on the Client. Ensure that the connecting adapters IDs match the profile of the Client and the second Server.

On the VIOC:

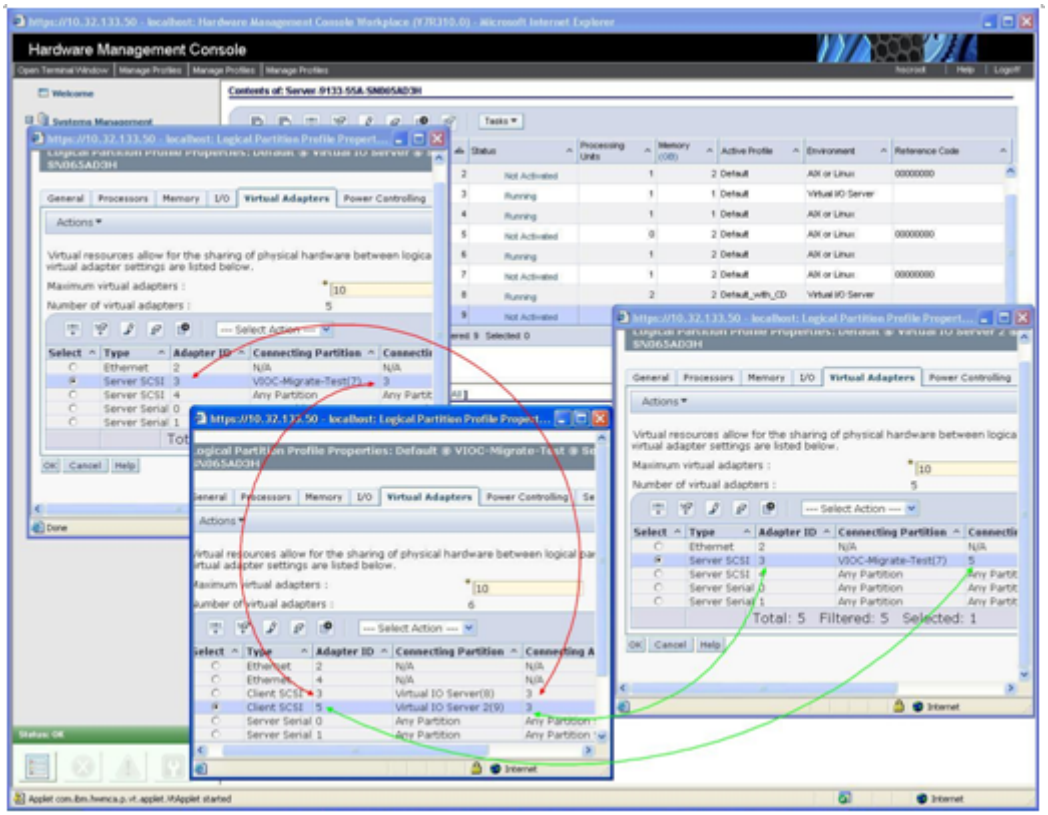
To add additional virtual adapters, follow these steps:

1. Shut down the Client partition.
2. Add a Virtual Ethernet adapter.

3. Add a Virtual SCSI adapter to its hardware profile through the HMC if it has not been created in the setup stage to provision for this Dual-VIOS setup.



4. Change the VLAN ID to use the same VLAN ID that was assigned to the second Virtual I/O Server.



- The small window on the left is the profile for Virtual I/O Server 1.
- The middle window is for the Virtual I/O Client.

- The window on the right belongs to Virtual I/O Server 2.

The Client SCSI Adapter must connect to the correct partition and the correct available Virtual SCSI Adapter ID. The Servers must be configured to allow the Client's respective Adapter ID to connect.

5. Re-activate the Client partition.
6. Enter **`lsdev -Ccadapter`** when it reboots. An additional Virtual Ethernet and Virtual SCSI Adapter has been added to the device list.

```
SGELIBM22/> lsdev -Ccadapter
ent0    Available         Virtual I/O Ethernet Adapter (l-lan)
ent1    Available       Virtual I/O Ethernet Adapter (l-lan)
...
fcs0    Available 06-08 FC Adapter
fcs1    Available 06-09 FC Adapter
...
vscsi0  Available         Virtual SCSI Client Adapter
vscsi1 Available       Virtual SCSI Client Adapter
```

7. Enter **lspath** and observe that a second path for hdisk2 has been added.

```
SGELIBM22/> lspath
Enabled hdisk2 vscsi0
Enabled hdisk2 vscsi1
```

8. Change the health check interval to **20**. This enables automatic path failure detection. The default value (0) disables this functionality.

IMPORTANT

A reboot is required for this change to take effect.

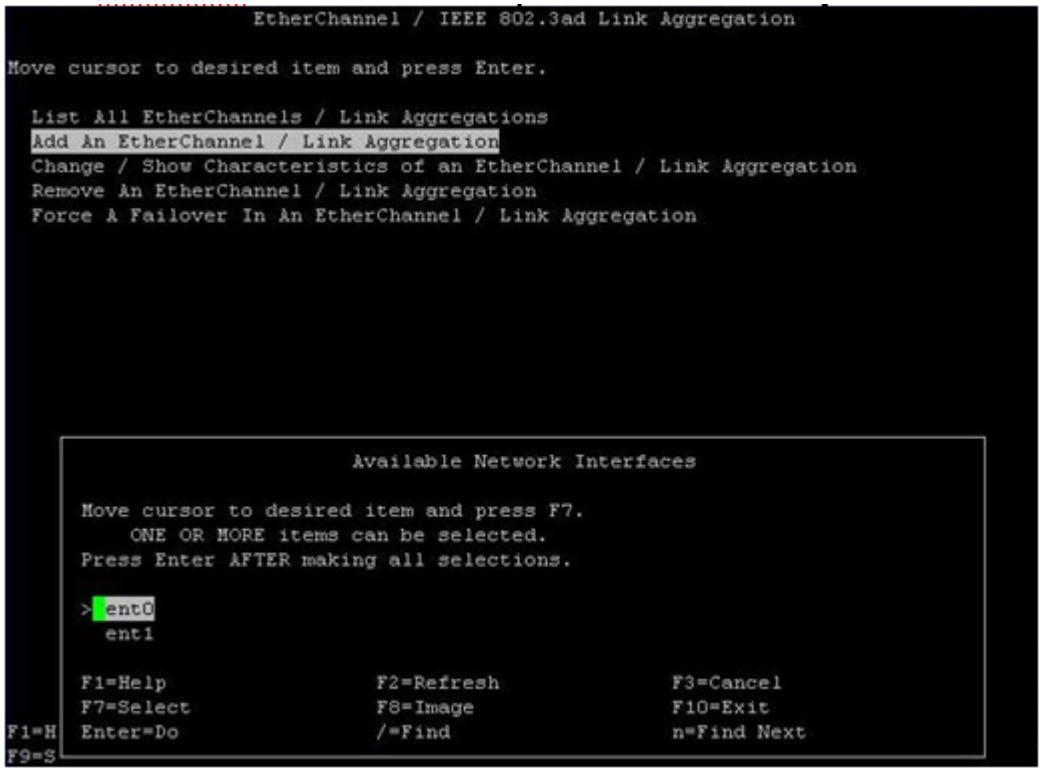
```
chdev -l hdisk2 -a hcheck_interval=20 -P
```

To aggregate Ethernet links:

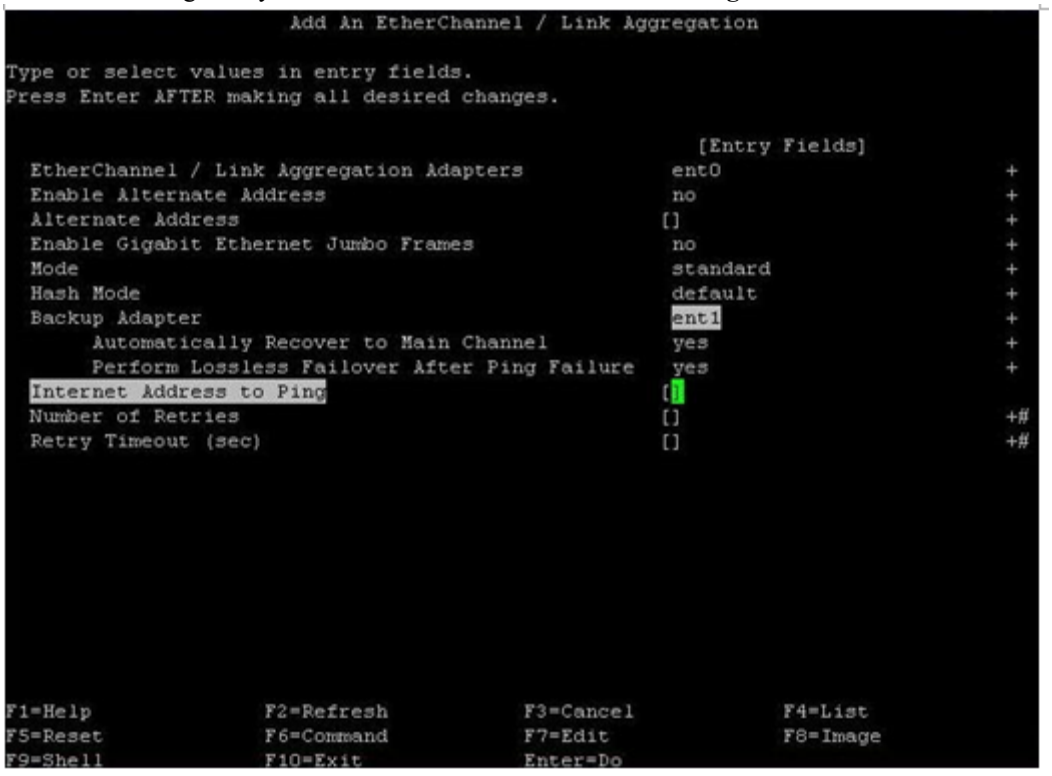
To enable redundancy for Ethernet links, aggregate the two virtual Ethernet adapters.

1. Use **`smitty ether channel`** to aggregate the two virtual Ethernet adapters.

2. Add an EtherChannel as shown in the following screenshot, and choose just one of the Virtual Adapters.



3. In the next menu, choose the other Virtual Ethernet Adapter as the backup adapter and enter the gateway IP address in the **Internet Address to Ping** field.



A new EtherChannel adapter has been created and your host will maintain connectivity through this new adapter. Therefore, users connecting through IP terminals would be disconnected.

4. Open a terminal through the HMC and continue working. The setup will complete soon.

```
SGELIBM22/> lsdev -Ccadapter
ent0   Available       Virtual I/O Ethernet Adapter (l-lan)
ent1   Available       Virtual I/O Ethernet Adapter (l-lan)
ent2   Available       EtherChannel / IEEE 802.3ad Link Aggregation
```

After setting the IP address on this new **ent2** adapter by using **smitty mktcpip**, the migration from a standalone LPAR to a Virtual I/O Client on dual Virtual I/O Servers has been successfully implemented.

IMPORTANT

Remember to test the functionality of failover by shutting down one of the Virtual I/O Servers.

Additional references

The reader may also be interested in the following documents available at Dell EMC Online Support:

- *PowerPath for AIX Installation and Administration Guide*

The following documents are available from IBM:

- Using the Virtual I/O Server

  http://publib.boulder.ibm.com/infocenter/eserver/v1r3s/topic/iphb1/iphb1.pdf

- Virtual I/O Server Commands Reference:

  http://publib.boulder.ibm.com/infocenter/eserver/v1r3s/topic/iphcg/iphcg.pdf

- POWER5 Virtualization: How to set up the IBM Virtual I/O Server:

  http://www.ibm.com/developerworks/eserver/library/es-aix-vioserver-v2

Creating a Fibre Channel boot device on the EMC system

This section describes requirements and procedures for booting the host from the Symmetrix, VNX series, or CLARiiON system.

Requirements and supported hardware and firmware

The following information provides general guidelines, requirements, and instructions for booting AIX from the Symmetrix, VNX series, or CLARiiON system. Refer to the *Dell EMC Simple Support Matrices* for supported hardware.

Operating system requirements

Minimum OS release levels are:

- AIX 5.1: Release Name 5100-01, VRMF 5.1.0.10, APAR IY21957
- AIX 5.2: Release Name 5200-01, VRMF 5.2.0.10, APAR IY44479
- AIX 5.3: Release Name 5300-00, VRMF 5.3.0.0
- AIX 6.1: Release Name 6100-00, 6100-00-01-0748

Recommended OS release levels are:

- AIX 5.1: Release Name 5100-05
- AIX 5.2: Release Name 5200-02
- AIX 5.3: Release Name 5300-00
- AIX 6.1: Release Name 6100-00-01-0748

Supported system firmware

You must obtain the most recent firmware level for the intended AIX server system from IBM. The website for the microcode updates is:

`http://techsupport/services.IBM.com/server/mdownload2/download.html`

EMC TimeFinder with Virtual I/O Server

This section contains the steps taken to migrate an AIX `rootvg` LPAR using Dell EMC TimeFinder®/Mirror, TimeFinder/Clone, and TimeFinder/Snap.

The migrated `rootvg` is then used to SAN boot into a Virtual I/O client. An example of such a migration is when a customer has requirements to test out an application patch on a production machine. Using TimeFinder /Snap copy services, the system administrator is able to create another replicated host for testing with minimal impact to the existing environment. Each of the copy services has their strengths and weaknesses. Refer to the appropriate TimeFinder Guide at Dell EMC Online Support to evaluate the best method for your environment.

The following sections provide more information:

- "Migration using TimeFinder copy services on Virtual I/O Server with MPIO or PowerPath" on page 148
- "Migration using TimeFinder/Clone to a STD device with copy on access" on page 152
- "Migration using TimeFinder /Clone to a STD device without copy on access " on page 155
- "Migration using TimeFinder/Clone to a target BCV device" on page 155
- "Migration using TimeFinder /Snap " on page 158

Migration using TimeFinder copy services on Virtual I/O Server with MPIO or PowerPath

This section provides what equipment is needed and the steps to take to migrate using TimeFinder/Mirror to a target BCV device.

Equipment setup

The following equipment is needed:

- 1 x VIOS with MPIO or PowerPath
- 1 x LPAR (`rootvg` boot from SAN. Symmetrix is used in the example.)
  - SGELIBM12
- 1 x VIOC
  - SGELIBM13
- 1 x AIX host with Solutions Enabler to run TimeFinder commands
  - SGELIBM11

Migration steps

To migrate using TimeFinder/Mirror to a target BCV device:

1. Find a BCV device that is the same size as the STD device which the LPAR is booting from.

   For this example, LPAR#1 is booting off a Symmetrix Meta LUN 06AE.

```
(0)SGELIBM12/> lsvg -p rootvg
rootvg:
PV_NAME           PV STATE          TOTAL PPs   FREE PPs    FREE DISTRIBUTION
hdiskpower162     active            359         203         71..08..00..52..72
```

```
(0)SGELIBM12/> powermt display dev=162
Pseudo name=hdiskpower162
Symmetrix ID=000290102606
Logical device ID=06AE
state=alive; policy=SymmOpt; priority=0; queued-IOs=0
==============================================================================
---------------- Host ---------------   - Stor -   -- I/O Path -  -- Stats ---
###  HW Path                I/O Paths    Interf.   Mode    State  Q-IOs Errors
==============================================================================
   1 fscsi1                   hdisk328  FA 14bA   active  alive      0      0
   0 fscsi0                   hdisk3    FA  3bA   active  alive      0      0
```

2. A Symmetrix BCV device 06C1 is chosen to be used in the TimeFinder session. Using a host with SE, set up the TimeFinder session using the following commands:

```
(0)SGELIBM11/> symdg create TF_M
 (0)SGELIBM11/> symld -g TF_M add dev 6ae
 (0)SGELIBM11/> symbcv -g TF_M associate dev 6c1
 (0)SGELIBM11/> symdg show TF_M

 Group Name:  TF_M

    Group Type                                : REGULAR
    Device Group in GNS                       : No
    Valid                                     : Yes
    Symmetrix ID                              : 000290102606
    Group Creation Time                       : Wed Jun 11 16:16:50 2008
    Vendor ID                                 : EMC Corp
    Application ID                            : SYMCLI

    Number of STD Devices in Group            :    1
    Number of Associated GK's                 :    0
    Number of Locally-associated BCV's        :    1
    Number of Locally-associated VDEV's       :    0
    Number of Locally-associated TGT's        :    0
    Number of Remotely-associated VDEV's(STD RDF):    0
    Number of Remotely-associated BCV's (STD RDF):    0
    Number of Remotely-associated TGT's(TGT RDF) :    0
    Number of Remotely-associated BCV's (BCV RDF):    0
    Number of Remotely-assoc'd RBCV's (RBCV RDF) :    0
    Number of Remotely-assoc'd BCV's (Hop-2 BCV) :    0
    Number of Remotely-assoc'd VDEV's(Hop-2 VDEV):    0
    Number of Remotely-assoc'd TGT's (Hop-2 TGT) :    0

    Standard (STD) Devices (1):
        {
        --------------------------------------------------------------------
                                                    Sym                Cap
        LdevName              PdevName              Dev  Att. Sts      (MB)
        --------------------------------------------------------------------
        DEV001                N/A                   06AE (M)   RW     23025
        }

    BCV Devices Locally-associated (1):
        {
        --------------------------------------------------------------------
                                                    Sym                Cap
        LdevName              PdevName              Dev  Att. Sts      (MB)
        --------------------------------------------------------------------
        BCV001                N/A                   06C1 (M)   RW     23025
        }
```

3. Start the TimeFinder mirror session and monitor the progress of the session.

```
(0)SGELIBM11/> symmir -g TF_M -full establish
```

```
Execute 'Full Establish' operation for device group
'TF_M' (y/[n]) ? y

'Full Establish' operation execution is in progress for
device group 'TF_M'. Please wait...

'Full Establish' operation successfully initiated for device group
'TF_M'.

(0)SGELIBM11/> symmir -g TF_M query


Device Group (DG) Name: TF_M
DG's Type             : REGULAR
DG's Symmetrix ID     : 000290102606


     Standard Device                   BCV Device                   State
--------------------------- ------------------------------------ ------------
                      Inv.                                 Inv.
Logical        Sym  Tracks Logical              Sym       Tracks STD <=> BCV
--------------------------- ------------------------------------ ------------

DEV001         06AE      0 BCV001              06C1 *         0 Synchronized

Total              -------                                -------
  Track(s)               0                                     0
  MB(s)                0.0                                   0.0

Legend:


(*): The paired BCV device is associated with this group.



Once the session shows synchronized, split the BCV.

(0)SGELIBM11/> symmir -g TF_M split

Execute 'Split' operation for device group
'TF_M' (y/[n]) ? y

'Split' operation execution is in progress for
device group 'TF_M'. Please wait...

'Split' operation successfully executed for device group
'TF_M'.

(0)SGELIBM11/> symmir -g TF_M query


Device Group (DG) Name: TF_M
DG's Type             : REGULAR
DG's Symmetrix ID     : 000290102606


     Standard Device                   BCV Device                   State
--------------------------- ------------------------------------ -----------
                      Inv.                                 Inv.
Logical        Sym  Tracks Logical              Sym       Tracks STD <=> BCV
--------------------------- ------------------------------------ -----------

DEV001         06AE      0 BCV001              06C1 *         0 Split
```

```
Total              -------                                -------
  Track(s)               0                                      0
  MB(s)                0.0                                    0.0

Legend:


(*): The paired BCV device is associated with this group.
```

4. Ensure that the VIOS is configured to see the TimeFinder BCV device.

   By default, the TimeFinder BCV device will be in a “Defined state”. To make it permanently available, run the script bcvfcpavail.sh in `/usr/lpp/EMC/Symmetrix/bin/bcvfcp.tar` and then run `cfgmgr` on VIOS. Use the `inq` utility to correctly identify which BCV LUN to use. Use the 5th to 7th digit of the serial number of the Symmetrix device to identify the correct BCV device. You might need to run **`powermt config`** after **`cfgmgr`** when your VIOS is using PowerPath. It should return as a `/dev/rhdiskpower` device from the inq output.

```
# lsdev -Ccdisk
hdisk0  Available 05-08-02     EMC Symmetrix FCP MPIO Raid1
hdisk1  Defined   05-08-02     EMC Symmetrix FCP MPIO Raid1 TimeFinder
hdisk2  Defined   05-08-02     EMC Symmetrix FCP MPIO Raid1 TimeFinder
hdisk3  Defined   05-08-02     EMC Symmetrix FCP MPIO Raid1 TimeFinder
hdisk4  Defined   05-09-02     EMC Symmetrix FCP MPIO Raid1 TimeFinder
hdisk12 Available 06-08-00-5,0 16 Bit LVD SCSI Disk Drive
hdisk13 Available 06-08-00-8,0 16 Bit LVD SCSI Disk Drive
hdisk5  Defined   05-09-02     EMC Symmetrix FCP MPIO Raid1 TimeFinder

# ./bcvfcpavail.sh
EMC Symmetrix BCV fcp devices changed to Available configuration state in the PdDv       ODM
object class.

Output of inq:
-------------------------------------------------------------------------
DEVICE            :VEND    :PROD            :REV   :SER NUM    :CAP(kb)
-------------------------------------------------------------------------
/dev/rhdiskpower0 :DGC     :RAID 5          :0428  :41000062   :    41943040
/dev/rhdiskpower1 :DGC     :VRAID           :0428  :420000EF   :    20971520
/dev/rhdisk0      :EMC     :SYMMETRIX       :5773  :0600042190 :       96000
/dev/rhdisk1      :EMC     :SYMMETRIX       :5773  :06006c5190 :    23577600
/dev/rhdisk2      :EMC     :SYMMETRIX       :5773  :06006c9190 :     5894400
/dev/rhdisk3      :EMC     :SYMMETRIX       :5773  :06006ca190 :     5894400
/dev/rhdisk4      :DGC     :LUNZ            :0428  :00000000   :      FAILED
/dev/rhdisk5      :EMC     :SYMMETRIX       :5773  :06006c1190 :    23577600
```

5. The vhost adapter is assumed to be already configured correctly between the VIOS and VIOC on the Hardware Management Console (HMC). In this example, the vhost adapter is vhost0. On the VIOS, create the virtual target device using the TimeFinder BCV LUN and the vhost adapter. Use the corresponding `hdiskpower` device instead, when using PowerPath on the VIOS.

```
$ mkvdev –vdev hdisk5 –vadapter vhost0 –dev sgelibm13_p1
$ lsmap -vadapter vhost0
SVSA            Physloc                                     Client Partition ID
--------------- ------------------------------------------- ------------------
vhost0          U9133.55A.065AD3H-V7-C3                     0x00000008

VTD                   sgelibm13_p1
Status                Available
LUN                   0x8100000000000000
Backing device        hdisk5
Physloc               U7311.D20.068766B-P1-C07-T1-W5006048452A75392-L5000000000000
```

6. Activate the VIOC and go into the SMS.
   - Select **5** to configure the boot options.
   - Select **1** to choose the boot device.
   - Select **5** to choose Hard Disk.
   - Select **1** to choose SCSI.
   - Select the correct vscsi adapter that was configured.
   - Select the correct virtual scsi disk.
7. The VIOC should boot up with the exact `rootvg` that was used in the TimeFinder session.

Migration using TimeFinder/Clone to a STD device with copy on access

This section provides what equipment is needed and the steps to take to migrate using TimeFinder/Clone to a STD device with copy on access.

Equipment setup

The following equipment is needed:

- 1 x VIOS with MPIO or PowerPath.
- 1 x LPAR (`rootvg` boot from SAN. Symmetrix is used in the example)
  - SGELIBM12
- 1 x VIOC
  - SGELIBM13
- 1 x AIX host with Solutions Enabler to run TimeFinder commands
  - SGELIBM11

Migration steps

To migration using TimeFinder/Clone to a STD device with copy on access:

1. Find a STD device that is the same size as the STD device which the LPAR is booting from.

   For this example, LPAR#1 is booting off a Symmetrix Meta LUN 06AE

```
(0)SGELIBM12/> lsvg -p rootvg
rootvg:
PV_NAME           PV STATE          TOTAL PPs   FREE PPs    FREE DISTRIBUTION
```

```
hdiskpower162     active            359         203         71..08..00..52..72

(0)SGELIBM12/> powermt display dev=162
Pseudo name=hdiskpower162
Symmetrix ID=000290102606
Logical device ID=06AE
state=alive; policy=SymmOpt; priority=0; queued-IOs=0
==============================================================================
--------------- Host ---------------   - Stor -   -- I/O Path -  -- Stats ---
###  HW Path                I/O Paths    Interf.   Mode    State  Q-IOs Errors
==============================================================================
   1 fscsi1                   hdisk328  FA 14bA   active  alive      0      0
   0 fscsi0                   hdisk3    FA  3bA   active  alive      0      0
```

2. A Symmetrix STD 06D1 is chosen to be used in the TimeFinder session. Using a host with SE, set up the TimeFinder session using the commands below:

```
(0)SGELIBM11/> symdg create TF_C
(0)SGELIBM11/> symld -g TF_C add dev 6ae
(0)SGELIBM11/> symld -g TF_C add dev 6d1
(0)SGELIBM11/> symdg show TF_C

Group Name:  TF_C

    Group Type                                  : REGULAR
    Device Group in GNS                         : No
    Valid                                       : Yes
    Symmetrix ID                                : 000290102606
    Group Creation Time                         : Wed Jun 11 16:16:50 2008
    Vendor ID                                   : EMC Corp
    Application ID                              : SYMCLI


    Number of STD Devices in Group              :    2
    Number of Associated GK's                   :    0
    Number of Locally-associated BCV's          :    0
    Number of Locally-associated VDEV's         :    0
    Number of Locally-associated TGT's          :    0
    Number of Remotely-associated VDEV's(STD RDF):    0
    Number of Remotely-associated BCV's (STD RDF):    0
    Number of Remotely-associated TGT's(TGT RDF) :    0
    Number of Remotely-associated BCV's (BCV RDF):    0
    Number of Remotely-assoc'd RBCV's (RBCV RDF) :    0
    Number of Remotely-assoc'd BCV's (Hop-2 BCV) :    0
    Number of Remotely-assoc'd VDEV's(Hop-2 VDEV):    0
    Number of Remotely-assoc'd TGT's (Hop-2 TGT) :    0

    Standard (STD) Devices (2):
        {
        --------------------------------------------------------------------
                                                    Sym                Cap
        LdevName              PdevName              Dev  Att. Sts      (MB)
        --------------------------------------------------------------------
        DEV001                N/A                   06AE (M)  RW      23025
        DEV002                N/A                   06D1 (M)  RW      23025
        }
```

3. Ensure that the VIOS is configured to see the STD target device. Run **cfgmgr** and use the **inq** utility to correctly identify which BCV LUN to use. Use the 5th to 7th digit of the serial number of the Symmetrix device to identify the correct STD device. You may need to run **powermt config** after **cfgmgr** when your VIOS is using PowerPath. It should return as a `/dev/rhdiskpower` device from the **inq** output.

Output of inq:

```
--------------------------------------------------------------------------------
DEVICE            :VEND    :PROD            :REV   :SER NUM    :CAP(kb)
--------------------------------------------------------------------------------
/dev/rhdiskpower0 :DGC     :RAID 5          :0428  :41000062   :    41943040
/dev/rhdiskpower1 :DGC     :VRAID           :0428  :420000EF   :    20971520
/dev/rhdiskpower2 :DGC     :RAID 5          :0326  :F300003A   :    31457280
/dev/rhdisk0      :EMC     :SYMMETRIX       :5773  :0600042190 :       96000
/dev/rhdisk1      :EMC     :SYMMETRIX       :5773  :06006c5190 :    23577600
/dev/rhdisk2      :EMC     :SYMMETRIX       :5773  :06006c9190 :     5894400
/dev/rhdisk3      :EMC     :SYMMETRIX       :5773  :06006ca190 :     5894400
/dev/rhdisk4      :DGC     :LUNZ            :0428  :00000000   :      FAILED
/dev/rhdisk5      :DGC     :LUNZ            :0326  :00000000   :      FAILED
/dev/rhdisk6      :DGC     :RAID 5          :0428  :41000062   :    41943040
/dev/rhdisk7      :DGC     :VRAID           :0428  :420000EF   :    20971520
/dev/rhdisk8      :EMC     :SYMMETRIX       :5773  :06006d1190 :    23577600
/dev/rhdisk9      :DGC     :RAID 5          :0428  :41000062   :    41943040
/dev/rhdisk10     :DGC     :VRAID           :0428  :420000EF   :    20971520
/dev/rhdisk11     :DGC     :LUNZ            :0428  :00000000   :      FAILED
/dev/rhdisk12     :IBM   H0:HUS151414VL3800 :S43C  :00A2975B   :   143374000
/dev/rhdisk13     :IBM   H0:HUS151414VL3800 :S43C  :00A29703   :   143374000
/dev/rhdisk14     :DGC     :RAID 5          :0326  :F300003A   :    31457280
```

4. Start the TimeFinder clone session and monitor the progress of the session

```
(0)SGELIBM11/> symclone -g TF_C create DEV001 sym ld DEV002

Execute 'Create' operation for device 'DEV001'
in device group 'TF_C' (y/[n]) ? y

'Create' operation execution is in progress for device 'DEV001'
paired with target device 'DEV002' in
device group 'TF_C'. Please wait...

'Create' operation successfully executed for device 'DEV001'
in group 'TF_C' paired with target device 'DEV002'.

(0)SGELIBM11/> symclone -g TF_C activate DEV001 sym ld DEV002

Execute 'Activate' operation for device 'DEV001'
in device group 'TF_C' (y/[n]) ? y

'Activate' operation execution is in progress for device 'DEV001'
paired with target device 'DEV002' in
device group 'TF_C'. Please wait...

'Activate' operation successfully executed for device 'DEV001'
in group 'TF_C' paired with target device 'DEV002'.
```

5. There is no need to wait for the clone group to fully synchronize. This is the difference between Mirror and Clone operations. The clone can be put to immediate use once the clone session is activated.
6. The vhost adapter is assumed to be already configured correctly between the VIOS and VIOC. In this example, the vhost adapter is vhost0. On the VIOS, create the virtual target device using the target LUN and the vhost adapter. Use the corresponding hdiskpower device instead, when using PowerPath on the VIOS.

```
$ mkvdev -vdev hdisk8 -vadapter vhost0 -dev sgelibm13_p2
sgelibm13_p2 Available



$ lsmap -vadapter vhost0
```

```
SVSA            Physloc                                     Client Partition ID
--------------- ------------------------------------------- ------------------
vhost0          U9133.55A.065AD3H-V7-C3                     0x00000008

VTD                   sgelibm13_p2
Status                Available
LUN                   0x8100000000000000
Backing device        hdisk8
Physloc               U7311.D20.068766B-P1-C07-T1-W5006048452A75392-L28000000000000
```

7. Activate the VIOC and go into the SMS.
   - Select **5** to configure the boot options.
   - Select **1** to choose the boot device.
   - Select **5** to choose Hard Disk.
   - Select **1** to choose SCSI
   - Select the correct vscsi adapter that was configured.
   - Select the correct virtual scsi disk.

   The VIOC should boot up with the exact `rootvg` that was used in the TimeFinder session.

Migration using TimeFinder /Clone to a STD device without copy on access

This section provides the equipment needed and the steps to take to migrate using TimeFinder/Clone to a STD device without copy on access.

Equipment setup The equipment needed is the same as outlined in "Equipment setup" on page 152.

Migration steps The steps are all the same as the procedure outlined in "Migration using TimeFinder/Clone to a STD device with copy on access" beginning on page 152. The only difference is to use the **`-copy`** argument in Step 7 when creating the TimeFinder/Clone session.

Without copy on access, a full copy of source LUN to target LUN will be done in the background.

Migration using TimeFinder/Clone to a target BCV device

This section provides the equipment needed and the steps to take to migrate using TimeFinder/Clone to a target BCV device.

Equipment setup The following equipment is needed:

- 1 x VIOS with MPIO or PowerPath.
- 1 x LPAR (`rootvg` boot from SAN. Symmetrix is used in the example)
  - SGELIBM12
- 1 x VIOC
- SGELIBM13

- 1 x AIX host with Solutions Enabler to run TimeFinder commands
  - SGELIBM11

Migration steps To migration using TimeFinder/Clone to a target BCV device:

1. Find a BCV device that is the same size as the STD device which the LPAR is booting from. For this example, LPAR#1 is booting off a Symmetrix Meta LUN 06AE.

```
(0)SGELIBM12/> lsvg -p rootvg
rootvg:
PV_NAME           PV STATE          TOTAL PPs   FREE PPs    FREE DISTRIBUTION
hdiskpower162     active      359     203     71..08..00..52..72

(0)SGELIBM12/> powermt display dev=162
Pseudo name=hdiskpower162
Symmetrix ID=000290102606
Logical device ID=06AE
state=alive; policy=SymmOpt; priority=0; queued-IOs=0
==============================================================================
--------------- Host ---------------   - Stor -   -- I/O Path -  -- Stats ---
###  HW Path                I/O Paths    Interf.   Mode    State  Q-IOs Errors
==============================================================================
   1 fscsi1                     hdisk328  FA 14bA   active  alive      0      0
 0 fscsi0          hdisk3   FA 3bA  active alive    0    0
```

2. A Symmetrix BCV 06C1 is chosen to be used in the TimeFinder session. Using a host with SE, set up the TimeFinder session using the following commands:

```
(0)SGELIBM11/> symdg create TF_C
(0)SGELIBM11/> symld -g TF_C add dev 6ae
(0)SGELIBM11/> symbcv -g TF_C associate dev 6c1
(0)SGELIBM11/> symdg show TF_C

Group Name:  TF_C

        Group Type                                   : REGULAR
    Device Group in GNS                          : No
    Valid                                        : Yes
    Symmetrix ID                                 : 000290102606
    Group Creation Time                          : Wed Jun 11 16:16:50 2008
    Vendor ID                                    : EMC Corp
    Application ID                               : SYMCLI

    Number of STD Devices in Group               :    1
    Number of Associated GK's                    :    0
    Number of Locally-associated BCV's           :    1
    Number of Locally-associated VDEV's          :    0
    Number of Locally-associated TGT's           :    0
    Number of Remotely-associated VDEV's(STD RDF):    0
    Number of Remotely-associated BCV's (STD RDF):    0
    Number of Remotely-associated TGT's(TGT RDF) :    0
    Number of Remotely-associated BCV's (BCV RDF):    0
    Number of Remotely-assoc'd RBCV's (RBCV RDF) :    0
    Number of Remotely-assoc'd BCV's (Hop-2 BCV) :    0
    Number of Remotely-assoc'd VDEV's(Hop-2 VDEV):    0
    Number of Remotely-assoc'd TGT's (Hop-2 TGT) :    0




Standard (STD) Devices (1):
        {
```

```
    ---------------------------------------------------------------
                                                Sym         Cap
    LdevName              PdevName              Dev  Att. Sts (MB)
    ---------------------------------------------------------------
    DEV001                N/A                   06AE (M)  RW 23025
    }

BCV Devices Locally-associated (1):
    {
    ---------------------------------------------------------------
                                                Sym         Cap
    LdevName              PdevName              Dev  Att. Sts (MB)
    ---------------------------------------------------------------
    BCV001                N/A                   06C1 (M)  RW 23025
    }
```

3. Ensure that the VIOS is configured to see the TimeFinder BCV device.

   By default, the TimeFinder BCV device will be in a "Defined state". To make it available, run the `bcvfcpavail.sh` in `/usr/lpp/EMC/Symmetrix/bin/bcvfcp.tar` and then cfgmgr on VIOS. Use the i**nq** utility to correctly identify which BCV LUN to use. Use the 5th to 7th digit of the serial number of the Symmetrix device to identify the correct BCV device. You may need to run **`powermt config`** after **`cfgmgr`** when your VIOS is using PowerPath. It should return as a `/dev/rhdiskpower` device from the inq output.

```
# lsdev -Ccdisk
hdisk0  Available 05-08-02     EMC Symmetrix FCP MPIO Raid1
hdisk1  Defined   05-08-02     EMC Symmetrix FCP MPIO Raid1 TimeFinder
hdisk2  Defined   05-08-02     EMC Symmetrix FCP MPIO Raid1 TimeFinder
hdisk3  Defined   05-08-02     EMC Symmetrix FCP MPIO Raid1 TimeFinder
hdisk4  Defined   05-09-02     EMC Symmetrix FCP MPIO Raid1 TimeFinder
hdisk12 Available 06-08-00-5,0 16 Bit LVD SCSI Disk Drive
hdisk13 Available 06-08-00-8,0 16 Bit LVD SCSI Disk Drive
hdisk5  Defined   05-09-02     EMC Symmetrix FCP MPIO Raid1 TimeFinder

# ./bcvfcpavail.sh
EMC Symmetrix BCV fcp devices changed to Available configuration state in the PdDv       ODM
object class.

Output of inq:
-------------------------------------------------------------------------
DEVICE            :VEND    :PROD            :REV   :SER NUM    :CAP(kb)
-------------------------------------------------------------------------
/dev/rhdiskpower0 :DGC     :RAID 5          :0428  :41000062   :    41943040
/dev/rhdiskpower1 :DGC     :VRAID           :0428  :420000EF   :    20971520
/dev/rhdisk0      :EMC     :SYMMETRIX       :5773  :0600042190 :       96000
/dev/rhdisk1      :EMC     :SYMMETRIX       :5773  :06006c5190 :    23577600
/dev/rhdisk2      :EMC     :SYMMETRIX       :5773  :06006c9190 :     5894400
/dev/rhdisk3      :EMC     :SYMMETRIX       :5773  :06006ca190 :     5894400
/dev/rhdisk4      :DGC     :LUNZ            :0428  :00000000   :      FAILED
/dev/rhdisk5      :EMC     :SYMMETRIX       :5773  :06006c1190 :    23577600
```

4. Start the TimeFinder clone session and monitor the progress of the session

```
0)SGELIBM11/> symclone -g TF_C create DEV001 sym ld BCV001 -copy

Execute 'Create' operation for device 'DEV001'
in device group 'TF_C' (y/[n]) ? y

'Create' operation execution is in progress for device 'DEV001'
paired with target device 'BCV001' in
device group 'TF_C'. Please wait...

'Create' operation successfully executed for device 'DEV001'
```

```
in group 'TF_C' paired with target device ' BCV001'.

(0)SGELIBM11/> symclone -g TF_C activate DEV001 sym ld BCV001

Execute 'Activate' operation for device 'DEV001'
in device group 'TF_C' (y/[n]) ? y

'Activate' operation execution is in progress for device 'DEV001'
paired with target device ' BCV001' in
device group 'TF_C'. Please wait...

'Activate' operation successfully executed for device 'DEV001'
in group 'TF_C' paired with target device ' BCV001'.
```

There is no need to wait for the clone group to fully synchronize. This is the difference between Mirror and Clone operations. The clone can be put to immediate use once the clone session is activated.

5. The vhost adapter is assumed to be configured correctly between the VIOS and VIOC. In this example, the vhost adapter is *vhost0*. On the VIOS, create the virtual target device using the target LUN and the vhost adapter. Use the corresponding *hdiskpower* device instead, when using PowerPath on the VIOS.

```
$ mkvdev -vdev hdisk5 -vadapter vhost0 -dev sgelibm13_p2
sgelibm13_p2 Available

$ lsmap -vadapter vhost0
SVSA            Physloc                                     Client Partition ID
--------------- ------------------------------------------- ------------------
vhost0          U9133.55A.065AD3H-V7-C3                     0x00000008

VTD                   sgelibm13_p2
Status                Available
LUN                   0x8100000000000000
Backing device        hdisk5
Physloc               U7311.D20.068766B-P1-C07-T1-W5006048452A75392-L28000000000000
```

6. Activate the VIOC and go into the SMS.
   - Select **5** to configure the boot options.
   - Select **1** to choose the boot device.
   - Select **5** to choose Hard Disk.
   - Select **1** to choose SCSI.
   - Select the correct vscsi adapter that was configured.
   - Select the correct virtual scsi disk.
7. The VIOC should boot up with the exact `rootvg` that was used in the TimeFinder session.

Migration using TimeFinder /Snap

This section provides the equipment needed and the steps to take to migrate using TimeFinder/Snap.

Equipment setup

The following equipment is needed:

- 1 x VIOS with MPIO or PowerPath.
- 1 x LPAR (`rootvg` boot from SAN. Symmetrix is used in the example.)
  - SGELIBM12
- 1 x VIOC
  - SGELIBM13
- 1 x AIX host with Solutions Enabler to run TimeFinder commands
  - SGELIBM11

Migration steps

To migrate using TimeFinder/Snap:

1. For TimeFinder/Snap, it is a requirement to have a VDEV that is the same config as the source STD device, such as A 4 x 5 GB LUNs STD Meta device will not work with a VDEV that is made up of a single contiguous 20 GB volume. You will need to have a VDEV that is also a 4 x 5 GB LUNs STD Meta device. Ensure that there is sufficient space in the SAVE pool.
2. Find a VDEV device that is the same size as the STD device which the LPAR is booting from. For this example, LPAR#1 is booting off a Symmetrix STD DEV 0853

```
(0)SGELIBM34/> lsvg -p rootvg
rootvg:
PV_NAME           PV STATE          TOTAL PPs   FREE PPs    FREE DISTRIBUTION
hdisk2            active            159         7           00..00..00..00..07
(0)SGELIBM34/> /usr/lpp/EMC/Symmetrix/bin/inq.aix32_51
Inquiry utility, Version V7.3-771 (Rev 0.0)      (SIL Version V6.3.0.0 (Edit Level 771)
Copyright (C) by EMC Corporation, all rights reserved.
For help type inq -h.

..

---------------------------------------------------------------------
DEVICE       :VEND    :PROD            :REV   :SER NUM    :CAP(kb)
---------------------------------------------------------------------
/dev/rhdisk1 :EMC     :SYMMETRIX       :5773  :6300112000 :       96000
/dev/rhdisk2 :EMC     :SYMMETRIX       :5773  :6300853000 :    20972160
```

A Symmetrix VDEV 0855 is chosen to be used in the TimeFinder session.

3. Using a host with SE, set up the TimeFinder session using the following commands:

```
# symdg create TF_S
# symld -g TF_S add dev 853
# symld -g TF_S add dev 855 -vdev
# symdg show TF_S


Group Name:  TF_S

    Group Type                                  : REGULAR
    Device Group in GNS                         : No
    Valid                                       : Yes
    Symmetrix ID                                : 000190300963
    Group Creation Time                         : Tue Jul 29 04:50:37 2008
    Vendor ID                                   : EMC Corp
    Application ID                              : SYMCLI

    Number of STD Devices in Group              :    1
    Number of Associated GK's                   :    0
    Number of Locally-associated BCV's          :    0
    Number of Locally-associated VDEV's         :    1
    Number of Locally-associated TGT's          :    0
    Number of Remotely-associated VDEV's(STD RDF):    0
    Number of Remotely-associated BCV's (STD RDF):    0
    Number of Remotely-associated TGT's(TGT RDF) :    0
    Number of Remotely-associated BCV's (BCV RDF):    0
    Number of Remotely-assoc'd RBCV's (RBCV RDF) :    0
    Number of Remotely-assoc'd BCV's (Hop-2 BCV) :    0
    Number of Remotely-assoc'd VDEV's(Hop-2 VDEV):    0
    Number of Remotely-assoc'd TGT's (Hop-2 TGT) :    0

    Standard (STD) Devices (1):
        {
        --------------------------------------------------------------------
                                                    Sym               Cap
        LdevName              PdevName              Dev  Att. Sts     (MB)
        --------------------------------------------------------------------
        DEV001                N/A                   0853      RW     20481
        }

    VDEV Devices Locally-associated (1):
        {
        --------------------------------------------------------------------
                                                    Sym               Cap
        LdevName              PdevName              Dev  Att. Sts     (MB)
        --------------------------------------------------------------------
        VDEV001               N/A                   0855      NR     20481
        }
```

4. Ensure that the VIOS is already configured to see the VDEV target device. Run **`cfgmgr`** and use the inq utility to correctly identify which BCV LUN to use. Use the 5th to 7th digit of the serial number of the Symmetrix device to identify the correct STD device. You may need to run **`powermt config`** after **`cfgmgr`** when your VIOS is using PowerPath. It should return as a `/dev/rhdiskpower` device from the `inq` output.

```
Output of inq:
Inquiry utility, Version V7.3-771 (Rev 0.0)      (SIL Version V6.3.0.0 (Edit Level 771)
Copyright (C) by EMC Corporation, all rights reserved.
For help type inq -h.

.....

-------------------------------------------------------------------------
DEVICE       :VEND    :PROD            :REV   :SER NUM    :CAP(kb)
```

```
---------------------------------------------------------------------
/dev/rhdisk0 :IBM     :HUS151473VL3800 :S430  :00ECD9F6   :    71687000
/dev/rhdisk1 :IBM     :ST373453LC      :C51C  :000B9743   :    71687000
/dev/rhdisk2 :EMC     :SYMMETRIX       :5773  :6300112000 :       96000
/dev/rhdisk3 :EMC     :SYMMETRIX       :5773  :6300823000 :    20974080
/dev/rhdisk4 :EMC     :SYMMETRIX       :5773  :6300855000 :    20972160
```

5. The vhost adapter is assumed to be configured correctly between the VIOS and VIOC. In this example, the vhost adapter is vhost1. On the VIOS, create the virtual target device using the target LUN and the vhost adapter.

```
$ mkvdev -vdev hdisk4 -vadapter vhost1 -dev sgelibm33_p1
sgelibm33_p1 Available
$ lsmap -vadapter vhost1
SVSA            Physloc                                     Client Partition ID
--------------- ------------------------------------------- ------------------
vhost1          U9133.55A.10AE20G-V5-C4                     0x00000000

VTD                   sgelibm33_p1
Status                Available
LUN                   0x8100000000000000
Backing device        hdisk4
Physloc               U787B.001.DNW69CA-P1-C3-T1-W50060482D5F0C8C1-L22000000000000

Ensure that the Symmetrix has sufficient space in the SAVE pools to hold data. For this case,
the default pool is used.

# symsnap list -savedevs

Symmetrix ID: 000190300963


                  S N A P   S A V E   D E V I C E S
-------------------------------------------------------------------------
      Device     SaveDevice        Total      Used      Free     Full
Sym   Emulation  Pool Name        Tracks    Tracks    Tracks      (%)
-------------------------------------------------------------------------
0839  FBA        DEFAULT_POOL      81930         0     81930        0
083A  FBA        DEFAULT_POOL      81930         0     81930        0
083B  FBA        DEFAULT_POOL      81930         0     81930        0
083C  FBA        DEFAULT_POOL      81930         0     81930        0
083D  FBA        DEFAULT_POOL      81930         0     81930        0
083E  FBA        DEFAULT_POOL      81930         0     81930        0
083F  FBA        DEFAULT_POOL      81930         0     81930        0
0840  FBA        DEFAULT_POOL      81930         0     81930        0
0841  FBA        DEFAULT_POOL      81930         0     81930        0
0842  FBA        DEFAULT_POOL      81930         0     81930        0
0843  FBA        DEFAULT_POOL      81930         0     81930        0
0844  FBA        DEFAULT_POOL      81930         0     81930        0
0845  FBA        DEFAULT_POOL      81930         0     81930        0
0846  FBA        DEFAULT_POOL      81930         0     81930        0
0847  FBA        DEFAULT_POOL      81930         0     81930        0

Total                          --------- --------- ---------     ----
  Tracks                         1228950         0   1228950        0
  MB(s)                          76809.4       0.0   76809.4

Create the symsnap session between the source and target device.

# symsnap create -g TF_S -v DEV001 vdev ld VDEV001

Execute 'Create' operation for device 'DEV001'
in device group 'TF_S' (y/[n]) ? y

'Create' operation execution is in progress for device 'DEV001'
paired with target device 'VDEV001' in
```

```
device group 'TF_S'. Please wait...


SELECTING the list of Source devices in the group:

  Device: 0853 [SELECTED]

SELECTING Target devices in the group:

  Device: 0855 [SELECTED]

PAIRING of Source and Target devices:

  Devices: 0853(S) - 0855(T) [PAIRED]

STARTING a Snap 'CREATE' operation.

The Snap 'CREATE' operation SUCCEEDED.

'Create' operation successfully executed for device 'DEV001'
in group 'TF_S' paired with target device 'VDEV001'.

Activate the symsnap session.

# symsnap activate -g TF_S -v DEV001 vdev ld VDEV001

Execute 'Activate' operation for device 'DEV001'
in device group 'TF_S' (y/[n]) ? y

'Activate' operation execution is in progress for device 'DEV001'
paired with target device 'VDEV001' in
device group 'TF_S'. Please wait...


SELECTING the list of Source devices in the group:

  Device: 0853 [SELECTED]

SELECTING Target devices in the group:

  Device: 0855 [SELECTED]

PAIRING of Source and Target devices:

  Devices: 0853(S) - 0855(T) [PAIRED]

STARTING a Snap 'ACTIVATE' operation.

The Snap 'ACTIVATE' operation SUCCEEDED.

'Activate' operation successfully executed for device 'DEV001'
in group 'TF_S' paired with target device 'VDEV001'.
```

6. Activate the VIOC and go into the SMS.
   - Select **5** to configure the boot options.
   - Select **1** to choose the boot device.
   - Select **5** to choose Hard Disk.
   - Select **1** to choose SCSI.

- Select the correct vscsi adapter that was configured.
- Select the correct virtual scsi disk.

The VIOC should boot up with the exact `rootvg` that was used in the Time Finder session.

7. As data changes on the source/target, storage will be allocated out of the save pool to hold the original snapshot data.

```
# symsnap list -savedevs

Symmetrix ID: 000190300963

                  S N A P   S A V E   D E V I C E S
--------------------------------------------------------------------
      Device     SaveDevice        Total      Used      Free     Full
Sym   Emulation  Pool Name        Tracks    Tracks    Tracks      (%)
--------------------------------------------------------------------
0839  FBA        DEFAULT_POOL      81930      2197     79733        2
083A  FBA        DEFAULT_POOL      81930      2197     79733        2
083B  FBA        DEFAULT_POOL      81930      2197     79733        2
083C  FBA        DEFAULT_POOL      81930      2197     79733        2
083D  FBA        DEFAULT_POOL      81930      2197     79733        2
083E  FBA        DEFAULT_POOL      81930      2197     79733        2
083F  FBA        DEFAULT_POOL      81930      2197     79733        2
0840  FBA        DEFAULT_POOL      81930      2197     79733        2
0841  FBA        DEFAULT_POOL      81930      2197     79733        2
0842  FBA        DEFAULT_POOL      81930      2197     79733        2
0843  FBA        DEFAULT_POOL      81930      2102     79828        2
0844  FBA        DEFAULT_POOL      81930      2028     79902        2
0845  FBA        DEFAULT_POOL      81930      2028     79902        2
0846  FBA        DEFAULT_POOL      81930      2114     79816        2
0847  FBA        DEFAULT_POOL      81930      2197     79733        2

Total                           --------- --------- ---------     ----
  Tracks                          1228950     32439   1196511        2
  MB(s)                           76809.4    2027.4   74781.9

# symsnap query -g TF_S

Device Group (DG) Name: TF_S
DG's Type             : REGULAR
DG's Symmetrix ID     : 000190300963

      Source Device          Target Device                  State     Copy
------------------------ ------------------- ---------- ------------ ----
               Protected                          Changed
Logical   Sym    Tracks    Logical   Sym     G    Tracks  SRC <=> TGT  (%)
------------------------ ------------------- ---------- ------------ ----

DEV001    0853     295251 VDEV001   0855    X      32439 CopyOnWrite    9

Total            --------                     ----------
  Track(s)         295251                          32439
  MB(s)           18453.2                         2027.4

Legend:

(G): X = The Target device is associated with this group,
     . = The Target device is not associated with this group.
```

CHAPTER 3

Migrating to the Dell EMC Supplied AIX ODM Package

This chapter is a reference point for migrating an existing IBM AIX 5.1/5.2/6.1 or 7.1/7.2 system running the Dell EMC old AIX ODM Package 5.3.*x*.*x* or 6.0.*x*.*x* to the AIX ODM Package that is supplied by Dell EMC.

Migration overview

This chapter is a reference point for migrating an existing IBM AIX 5.1/5.2/5.3/6.1 or 7.1/7.2 system running the Dell EMC old AIX ODM Package 5.3.0.x or 6.0.0.x to the Dell EMC AIX ODM Package that is supplied by Dell EMC.

Users of this chapter should have a working knowledge of AIX operating system commands and CLARiiON and/or Symmetrix storage arrays.

The migration process must be performed while the host is offline. This means that all applications must be offline and shut down, file systems unmounted, and volume groups varied off.

The latest ODM version:

- EMC.AIX.5.3.1.0.tar.Z

  Support for AIX 5.2/5.3/6.1 and VIOS

  Update only from EMC.AIX.5.1.0.0.tar.Z or higher

- EMC.AIX.6.0.0.5.tar.Z

  Support for AIX 7.1/7.2

  Update only from EMC.AIX.6.0.0.1.tar.Z or higher

Required software revision levels

The following list shows required software levels for performing the migration to the AIX ODM Package that is supplied by EMC.

Software marked with an asterisk (*) indicates that the required revision level must be installed prior to the migration:

- AIX Operating System version (for ODM version 5.3.x.x): *
  - 5.2.0.0
  - 5.3.0.0
  - 6.1.0.0
- AIX Operating System version (for ODM version 6.0.x.x): *
  - 7.1.0.0
  - 7.2.1.1
- Dell EMC software:
  - PowerPath Version 3.0.3 or higher (if applicable)
  - AIX ODM 5.3.0.0or 6.0.0.0 or higher*

Gathering host information

We recommend the following commands for gathering host prerequisite information and for host-specific configuration information. Before making any changes, collect as much information as possible to document the current host configuration.

- Display AIX operating system version:

  ```
  oslevel -s
  ```

- Display EMC AIX ODM revision:

  ```
  lslpp -L | grep EMC
  ```

- Display IBM Fibre Channel driver revision:

  ```
  lslpp -l devices.pci.df1000f7.com
  lslpp -l devices.pci.df1000f7.diag
  lslpp -l devices.pci.df1000f7.rte
  lslpp -l devices.pci.df1000f9.diag
  lslpp -l devices.pci.df1000f9.rte
  lslpp -l devices.pci.df1080f9.diag
  lslpp -l devices.pci.df1080f9.rte
  lslpp -l devices.fcp.disk.rte
  ```

- List hdisk entries:

  ```
  lsdev -Cc disk
  ```

- List hdiskpower entries:

  ```
  lsdev -Ct power
  ```

- Display PowerPath revision:

  ```
  lslpp -L | grep EMCpower
  ```

- List installed IBM native Fibre Channel HBAs:

  ```
  lsdev -Cc adapter |grep fcs
  ```

- List available configured IBM driver instances:

  ```
  lsdev -Cc driver | grep fscsi
  ```

- List the configured adapter, driver, and device instances and send the output to a file:

  ```
  lscfg
  ```

- List the hdisk PVIDs:

  ```
  lspv
  ```

- List volume group information:

  ```
  lsvg
  ```

Migration for nonboot on CLARiiON and Symmetrix arrays

Migration time will vary depending on the configuration size. Ensure that the customer has allocated an appropriate amount of time for the migration.

These steps assume that both the host operating system and array firmware are updated to correct levels. (Refer to "Required software revision levels" on page 167.)

1. If the Unisphere/Navisphere host agent is running, stop the agent:

   ```
   /etc/rc.agent stop
   ```

2. Verify that all CLARiiON and Symmetrix file systems are unmounted and all volume groups are offline. Stop and shut down any remaining applications that are running. Confirm that all volume groups are varied off:

   ```
   lsvg -o
   ```

3. Remove all EMC hdiskpower and hdisk devices by typing the following commands:

   ```
   powermt remove dev=all
   lsdev -Ct power -cdisk -F name | xargs -n1 rmdev –dl
   lsdev –Ct SYMM* -F name | xargs -n1 rmdev –dl
   lsdev –Ct CLAR* -F name | xargs -n1 rmdev –dl
   rmdev -dl powerpath0 (if applicable)
   savebase -v
   ```

4. Check the AIX ODM Package Version:

   ```
   lslpp -L | grep EMC
   ```

5. If the AIX ODM Package 5.3.x.x or 6.0.x.x is installed, then uninstall as follows:

   ```
   installp -u EMC.Symmetrix*
   installp -u EMC.CLARiiON*
   installp -u EMC.XtremIO*
   ```

   To download the AIX ODM Package Version 5.3.x.x or 6.0.x.x from ftp.emc.com:

   a. Type this command:

      ```
      ftp ftp.emc.com
      ```

   b. Log in with a username of anonymous and your e-mail address as a password.

   c. Type this command:

      ```
      cd /pub/elab/aix/ODM_DEFINITIONS
      ```

   d. Type this command:

      ```
      get EMC.AIX.5.1.0.X.tar.Z
      ```

   e. Type this command:

      ```
      quit
      ```

6. Install the AIX ODM Package Version 5.3.x.x or 6.0.x.x using the AIX **`installp`** command (**`smitty installp`**). Refer to the ODM Package README file for specific installation instructions.

7. Run the **`emc_cfgmgr`** command to ensure devices are configured correctly down all paths.

```
/usr/lpp/EMC/CLARiiON/bin/emc_cfgmgr
```

8. Confirm that all the EMC devices were configured down all the appropriate Fibre Channel paths as expected. At a minimum, type the following commands:

```
lsdev -Cc disk
/usr/lpp/EMC/CLARiiON/bin/inq.aix32_51
lspv
```

9. If PowerPath is installed, confirm that all PowerPath devices are configured correctly by typing the following commands:

```
powermt config
powermt display dev=all
```

10. Restart the system by typing:

```
shutdown -Fr
```

At this point, the migration process is complete. The customer may start their applications.

Migration when booting from CLARiiON storage arrays

This section covers the migration steps when `rootvg` is built on CLARiiON AIX MPIO hdisk and PowerPath hdiskpower devices. As with any procedure involving rootvg, a mksysb should be performed to back up any critical data.

Prerequisites

Before migrating the host, the zoning configuration of the host must be reconfigured so that there is only one path from the host to the array. For example, the fcs0 adapter is zoned to see one SPA port on the CLARiiON array and the SPA is the current owner of the boot LUN. This reconfiguration is required because there is a point in the migration when only one remaining path is left enabled between the host and the array.

After the zoning is reconfigured, remove all missing paths (MPIO SAN boot host) and defined hdisks (PowerPath SAN boot host) before proceeding.

The steps described in "Migration steps when booting from CLARiiON AIX MPIO hdisk or PowerPath hdiskpower devices" on page 171 are for a host where fcs0 is zoned to one SPA port and fcs1 is zoned to one SPB port. In such a case, there is no need to modify the existing zone because disabling one of the HBA adapter's SAN switch ports will only leave the host with one path to the CLARiiON array.

Migration steps when booting from CLARiiON AIX MPIO hdisk or PowerPath hdiskpower devices

A `boot_change_v4.sh` utility script was developed to enable the removal of existing CLARiiON hdisk devices by changing the CLARiiON ODM de-installation requirements.

1. Download the Dell EMC-supplied `boot_change_v4.tar` package from the ftp server ftp.emc.com.
   a. Type the following command:

      `ftp ftp.emc.com`

   b. Log in with a username of anonymous and your email address as a password.
   c. Type the following command:

      `cd /pub/elab/aix/ODM_DEFINITIONS`

   d. Type the following command:

      `get boot_change_v4.tar`

   e. Type the following command:

      `quit`

2. Extract the files from the boot_change_v4.tar package by typing:

   `tar -xvf boot_change_4.tar`

   The following files will be extracted into the user's existing directory path:

- **boot_change.README**
- **boot_change_v4.sh**
- **osfcpdisk.add**
- **osfcpmpiodisk.add**
- **osscsidisk.add**

3. Type **pprootdev off** on the host and reboot if the host is SAN booted on the hdiskpower device.
4. Uninstall PowerPath.
5. Use the `script boot_change_v4.sh` to convert the existing CLARiiON ODM definitions:

   ./boot_change_v4.sh

   The script will request input from the user. Enter **S** for Symmetrix devices and **C** for CLARiiON devices. A message similar to the following will display for each hdisk device that was successfully converted.

   - For PowerPath booted hosts:

```
CLARiiON FCP device hdisk2 was converted to Other FC SCSI Disk Drive
CLARiiON FCP device hdisk3 was converted to Other FC SCSI Disk Drive
CLARiiON FCP device hdisk4 was converted to Other FC SCSI Disk Drive
CLARiiON FCP device hdisk5 was converted to Other FC SCSI Disk Drive
```

   - For MPIO booted hosts:

```
CLARiiON FCP MPIO device hdisk2 was converted to Other FC MPIO SCSI Disk Drive
CLARiiON FCP MPIO device hdisk3 was converted to Other FC MPIO SCSI Disk Drive
CLARiiON FCP MPIO device hdisk4 was converted to Other FC MPIO SCSI Disk Drive
CLARiiON FCP MPIO device hdisk5 was converted to Other FC MPIO SCSI Disk Drive
```

6. Remove all converted hdisks that are not part of `rootvg` using **rmdev -dl hdisk*X*.** A **hdisk2 deleted** message will display for each hdisk device that is successfully removed.
7. Uninstall the CLARiiON ODM filesets by typing the following command:

   installp -u EMC.CLARiiON.*

8. Download the AIX ODM Package Version 5.3.x.x or 6.0.x.x from the ftp server ftp.emc.com.

   a. Type the following command:

      ftp ftp.emc.com

   b. Log in with a username of anonymous and your email address as a password.

   c. Type the following command:

      cd /pub/elab/aix/ODM_DEFINITIONS

   d. Type the following command:

      get EMC.AIX.5.3.0.X.tar.Z

   e. Type the following command:

      quit

9. Place Dell EMC Supplied AIX ODM Package Version 5.3.0.X into the **/usr/sys/inst.images** directory for installation. Uncompress and untar the package. Type the following:

   ```
   cd /usr/sys/inst.images
   uncompress EMC.AIX.5.3.0.X.tar.Z
   tar -xvf EMC.AIX.5.3.0.X.tar
   ```

10. Install the desired CLARiiON ODM filesets.
11. Block the SAN switch port of the host HBA that connects to the passive path to the CLARiiON array for the boot LUN. To find out the passive path, check the boot LUN ownership in `CLARiiON Unisphere/Navisphere Manager` under properties of the boot LUN.
12. If the host is an MPIO SAN booted host, remove the passive path to the boot LUN by typing the following path:

    ```
    rmpath -l hdiskX -p fscsiY -d.
    ```

13. Configure the hdisk devices as CLARiiON by typing the following command. All devices will be updated except `rootvg`. This will be updated after the system reboot is performed:

    ```
    /usr/lpp/EMC/CLARiiON/bin/emc_cfgmgr
    ```

14. Type the following command to reboot the host so that the current `rootvg hdisk` devices will be reconfigured as CLARiiON devices:

    ```
    shutdown -Fr
    ```

15. Type the following command to check the hdisks to ensure they are configured as CLARiiON devices:

    ```
    lsdev -Cc disk
    ```

    - For PowerPath booted hosts:

```
hdisk2 Available 1n-08-01 EMC CLARiiON FCP Raid1
hdisk3 Available 1n-08-01 EMC CLARiiON FCP Raid1
hdisk4 Available 1n-08-01 EMC CLARiiON FCP Raid1
hdisk5 Available 1n-08-01 EMC CLARiiON FCP Raid1
```

    - For MPIO booted hosts:

```
hdisk2 Available 1n-08-01 EMC CLARiiON FCP MPIO Raid1
hdisk3 Available 1n-08-01 EMC CLARiiON FCP MPIO Raid1
hdisk4 Available 1n-08-01 EMC CLARiiON FCP MPIO Raid1
hdisk5 Available 1n-08-01 EMC CLARiiON FCP MPIO Raid1
```

16. Enable the previously disabled SAN switch port and revert back to the original zoning configuration if required.
17. Type the following command to configure the new paths.:

    ```
    cfgmgr -v
    ```

18. Re-install and configure the PowerPath SAN boot if required.
19. If MPIO ODM is installed, check that all paths are enabled on the hdisks. If the paths are not enabled, reboot the host again. Remove any defined hdisks if host was converted from PowerPath to MPIO SAN boot.

At this point, the migration process is complete and you can start your applications.

Migration when booting from Symmetrix storage arrays

This section covers the migration steps when `rootvg` is built on Symmetrix AIX MPIO hdisk and PowerPath hdiskpower devices. As with any procedure involving `rootvg`, an `mksysb` should be performed to back up any critical data and all additional volume groups should be varied offline.

Migration steps when booting from Symmetrix AIX MPIO hdiskpower devices

You can use a `boot_change_v4.sh` utility script to remove existing Symmetrix hdisk devices by changing the Symmetrix ODM deinstallation requirements.

1. Download the `boot_change_v4.tar` package from the ftp server ftp.emc.com.
   a. Type the following command:

      `ftp ftp.emc.com`

   b. Log in with a username of anonymous and your email address as a password.
   c. Type the following command:

      `cd /pub/elab/aix/ODM_DEFINITIONS`

   d. Type the following command:

      `get boot_change_v4.tar`

   e. Type the following command:

      `quit`

2. Extract the files from the boot_change_v4.tar package by typing:

   `tar -xvf boot_change_v4.tar`

   The following files will be extracted into the user's existing directory path:

   - `boot_change.README`
   - `boot_change_v4.sh`
   - `osfcpdisk.add`
   - `osfcpmpiodisk.add`
   - `osscsidisk.add`

3. Run **`pprootdev off`** on the host and reboot with rootvg on the hdisk if the host is SAN booted on the hdiskpower device.
4. Use the script boot_change_v4.sh to convert the existing Symmetrix ODM definitions:

   `./boot_change_v4.sh`

   The script will request input from the user. Enter **S** for Symmetrix devices and **C** for CLARiiON devices. A message similar to the following will display for each hdisk device that was successfully converted.

   - For PowerPath booted hosts:

```
SYMMETRIX FCP device hdisk2 was converted to Other FC SCSI Disk Drive
SYMMETRIX FCP device hdisk3 was converted to Other FC SCSI Disk Drive
SYMMETRIX FCP device hdisk4 was converted to Other FC SCSI Disk Drive
SYMMETRIX FCP device hdisk5 was converted to Other FC SCSI Disk Drive
```

- For MPIO booted hosts:

```
SYMMETRIX FCP MPIO device hdisk2 was converted to Other FC MPIO SCSI Disk Drive
SYMMETRIX FCP MPIO device hdisk3 was converted to Other FC MPIO SCSI Disk Drive
SYMMETRIX FCP MPIO device hdisk4 was converted to Other FC MPIO SCSI Disk Drive
SYMMETRIX FCP MPIO device hdisk5 was converted to Other FC MPIO SCSI Disk Drive
```

5. Remove all converted hdisks that are not part of rootvg using **rmdev -dl hdisk*X*.** A message similar to the following will display for each hdisk device that was successfully removed:

   ```
   hdisk2 deleted
   hdisk3 deleted
   hdisk4 deleted
   hdisk5 deleted
   ```

6. Uninstall the Symmetrix ODM filesets by typing the following command:

   installp -u Symmetrix.*

7. Download the AIX ODM Package Version 5.3.0.x or 6.0.x.x from the ftp server ftp.emc.com.

   a. Type the following command:

      ftp ftp.emc.com

   b. Log in with a username of anonymous and your email address as a password.

   c. Type the following command:

      cd /pub/elab/aix/ODM_DEFINITIONS

   d. Type the following command:

      get EMC.AIX.5.1.0.X.tar.Z

   e. Type the following command:

      quit

8. Place the AIX ODM Package Version 5.1.0.X into the **/usr/sys/inst.images** directory for installation. Uncompress and untar the package. Type the following:

   ```
   cd /usr/sys/inst.images
   uncompress EMC.AIX.5.1.0.X.tar.Z
   tar -xvf EMC.AIX.5.10.X.tar
   ```

9. Install the Symmetrix ODM filesets.

10. Configure the hdisk devices as CLARiiON by typing the following command. All devices will be updated except `rootvg`. This will be updated after the system reboot is performed:

    /usr/lpp/EMC/Symmetrix/bin/emc_cfgmgr

11. Reboot the host so that the current rootvg hdisk devices will be reconfigured as CLARiiON devices by typing:

    shutdown -Fr

12. Check the hdisks to be sure they are configured as CLARiiON devices by typing:

    lsdev -Cc disk

- For PowerPath booted hosts

```
hdisk2 Available 1n-08-01 EMC Symmetrix FCP Raid1
hdisk3 Available 1n-08-01 EMC Symmetrix FCP Raid1
hdisk4 Available 1n-08-01 EMC Symmetrix FCP Raid1
hdisk5 Available 1n-08-01 EMC Symmetrix FCP Raid1
```

- For MPIO booted hosts

```
hdisk2 Available 1n-08-01 EMC Symmetrix FCP MPIO Raid1
hdisk3 Available 1n-08-01 EMC Symmetrix FCP MPIO Raid1
hdisk4 Available 1n-08-01 EMC Symmetrix FCP MPIO Raid1
hdisk5 Available 1n-08-01 EMC Symmetrix FCP MPIO Raid1
```

13. Configure thePowerPath SAN boot if required.

At this point, the migration process is complete and you can start your applications.

Migration steps when booting from Symmetrix PowerPath hdiskpower devices

1. Type the following command to turn off multipathing to the root device:

   pprootdev off

2. Reboot the host so that the boot device is no longer under PowerPath control. Type:

   shutdown -Fr

3. When the host comes back up, stop the Unisphere/Navisphere host agent if it is running by typing:

   /etc/rc.agent stop

4. Remove any ghost disk devices that were created. First clean up PowerPath by running the **powermt** check and answer ***yes*** to confirm the removal of any dead paths by typing:

   powermt check

   ```
   Warning: Symmetrix device path hdisk4 is currently
   dead. Do you want to remove it (y/n/a/q)? y
   ```

5. List and remove any Symmetrix hdisk devices that are in the “Defined” state by typing:

   lsdev –CtSYMM* -SD –F name | xargs –n1 rmdev -dl

6. Remove all configured hdisk and hdiskpower devices that are not part of `rootvg` by typing:

   powermt display

```
Symmetrix logical device count=50
================================================================================
----- Host Bus Adapters --------- ------ I/O Paths ----- ------ Stats ------
### HW Path Summary Total Dead IO/Sec Q-IOs Errors
================================================================================
0 fscsi0 optimal 50 0 - 0 0
1 fscsi1 optimal 50 0 - 0 0
```

```
powermt remove hba=0
powermt remove hba=1

Warning: removing last active path to volume 003D of Symmetrix
storage array 000187900087.
Proceed? [y|n|a|q] a
```

```
powermt remove dev=all
lsdev -Ct power -cdisk -F name | xargs -n1 rmdev -dl

hdiskpower0 deleted
hdiskpower1 deleted
hdiskpower2 deleted
rmdev -dl powerpath0
powerpath0 deleted
savebase -v
```

7. Download the `boot_change_v4.tar` package from the ftp server ftp.emc.com:
   a. Type the following command:

      `ftp ftp.emc.com`

   b. Log in with a username of anonymous and your e-mail address as a password.
   c. Type the following command:

      `cd /pub/elab/aix/ODM_DEFINITIONS`

   d. Type the following command:

      `get boot_change_v4.tar`

   e. Type the following command:

      `quit`

8. Type the following command to extract the files from the `boot_change.tar` package:

   `tar -xvf boot_change.tar`

   The following files will be extracted into the existing directory path:

   ```
   boot_change.README
   boot_change_v4.sh
   osfcpdisk.add
   osfcpmpiodisk.add
   osscsidisk.add
   ```

9. Convert the existing Symmetrix hdisk ODM definitions to “Other FC SCSI Disk Drive”. Type:

   `./boot_change_v4.sh`

   A message, similar to the following, will be displayed for each hdisk device that was successfully converted:

```
SYMMETRIX FCP device hdisk2 was converted to Other FC SCSI Disk Drive
SYMMETRIX FCP device hdisk3 was converted to Other FC SCSI Disk Drive
SYMMETRIX FCP device hdisk4 was converted to Other FC SCSI Disk Drive
SYMMETRIX FCP device hdisk5 was converted to Other FC SCSI Disk Drive
```

10. Remove all converted hdisks which are not part of rootvg by typing:

    `lsdev -Ct osdisk -F name | xargs -n1 rmdev -dl`

    A message similar to the following will be displayed for each hdisk device that was successfully removed:

    ```
    hdisk2 deleted
    hdisk3 deleted
    hdisk4 deleted
    hdisk5 deleted
    ```

11. Uninstall the Symmetrix ODM filesets by typing the following command:

    ```
    installp -u Symmetrix.*
    ```

12. Download the AIX ODM Package Version 5.3.x.x or 6.0.x.x from the EMC ftp server ftp.emc.com:

    a. Type the following command:

    ```
    ftp ftp.emc.com
    ```

    b. Log in with a username of anonymous and your e-mail address as a password.

    c. Type the following command:

    ```
    cd /pub/elab/aix/ODM_DEFINITIONS
    ```

    d. Type the following command:

    ```
    get EMC.AIX.5.3.0.X.tar.Z
    ```

    e. Type the following command:

    ```
    quit
    ```

13. Place the EMC AIX ODM Package Version 5.1.0.X into the **/usr/sys/inst.images** directory for installation. Uncompress and untar the package. Enter:

    ```
    cd /usr/sys/inst.images
    uncompress EMC.AIX.5.1.0.X.tar.Z
    tar –xvf EMC.AIX.5.1.0.X.tar
    ```

14. Install the Symmetrix ODM filesets by typing the following command:

    ```
    installp -acgXd . EMC.Symmetrix.aix.rte
    EMC.Symmetrix.fcp.rte
    ```

15. Configure the hdisk devices as Symmetrix by typing the following command. All devices will be updated except rootvg. This will be updated after the system reboot is performed in step 16:

    ```
    /usr/lpp/EMC/Symmetrix/bin/emc_cfgmgr
    ```

16. Reboot the host so that the current hdisk rootvg devices will be reconfigured as Symmetrix devices by typing:

    ```
    shutdown -Fr
    ```

17. Configure PowerPath devices by typing:

    ```
    powermt config
    ```

18. Enable PowerPath as the boot device by typing:

    ```
    pprootdev on
    ```

19. Reboot the host so that PowerPath will now control the Symmetrix boot device by typing:

    ```
    shutdown -Fr
    ```

At this point, the migration process is complete. The customer may start their applications.

Migration when booting from XtremIO storage arrays

This section provides the migration steps when `rootvg` is built on XtremIO AIX MPIO hdisk and PowerPath hdiskpower devices. As this procedure involves `rootvg`, you must perform an `mksysb` backup of all critical data and verify all additional volume groups offline.

Migration steps when booting from XtremIO AIX MPIO hdiskpower devices

You can use a `boot_change_v4.sh` utility script to remove existing XtremIO hdisk devices by changing the XtremIO ODM deinstallation requirements.

1. Complete the following steps to download the **`boot_change_v4.tar`** package from the ftp server:

   Note: If your system is running ODM version 6.1.1.1., download the `boot_change_v4_xio_6111.tar` package

   a. Run the following command to go to the ftp server:

      `ftp ftp.emc.com`

   b. Log in with an anonymous username and your email ID as password.

   c. Run the following command to go to the package location:

      `cd /pub/elab/aix/ODM_DEFINITIONS`

   d. Run the following command to enter the binary mode:

      `binary`

   e. Run one of the following commands to download the package:

      `get boot_change_v4.tar`

      or

      `get boot_change_v4_xio_6111.tar` (when the installed ODM version is 6.1.1.1)

   f. Run the following command to exit:

      `quit`

2. Run the following command to extract the files from the **`boot_change_v4.tar`** package:

   `tar -xvf boot_change_v4.tar`

   The following files will be extracted into the user's existing directory:

   - `boot_change.README`
   - `boot_change_v4.sh`
   - `osfcpdisk.add`
   - `osfcpmpiodisk.add`
   - `osscsidisk.add`

3. Use the following script to convert the existing XtremIO ODM definitions:

   `./boot_change_v4.sh`

   This script will request input from the user. Enter **X** for XtremIO devices. A message similar to the following example displays for each hdisk device that was successfully converted:

```
XtremIO FCP MPIO device hdisk2 was converted to OTHER FC MPIO SCSI DISK DRIVE
XtremIO FCP MPIO device hdisk3 was converted to OTHER FC MPIO SCSI DISK DRIVE
XtremIO FCP MPIO device hdisk4 was converted to OTHER FC MPIO SCSI DISK DRIVE
XtremIO FCP MPIO device hdisk5 was converted to OTHER FC MPIO SCSI DISK DRIVE
```

4. Run the following command to remove all the converted hdisks that are not part of `rootvg`:

   `lsdev -Ct mpioosdisk -F name | xargs -n1 rmdev -dl`

   A message similar to the following example displays for each hdisk device that was successfully removed:

   ```
   hdisk2 deleted
   hdisk3 deleted
   hdisk4 deleted
   hdisk5 deleted
   ```

5. Run the following command to uninstall the XtremIO ODM filesets:

   `installp -u EMC.XtremIO.aix.rte EMC.XtremIO.fcp.MPIO.rte`

6. Complete the following steps to download the AIX ODM Package Version 5.3.x.x or 6.x.x.x from the ftp server:

   a. Run the following command to go to the ftp server:

      `ftp ftp.emc.com`

   b. Log in with an anonymous username and your email ID as password.

   c. Run the following command to go to the AIX ODM package location:

      `cd /pub/elab/aix/ODM_DEFINITIONS`

   d. Run the following command to download the AIX ODM package:

      `get EMC.AIX.5.3.x.x.tar.Z`

   e. Run the following command to exit:

      `quit`

7. Place the AIX ODM Package Version 5.3.x.x in the **`cd /usr/sys/inst.images`** directory for installation.

8. Run the following command to uncompress and untar the package:

   ```
   uncompress EMC.AIX.5.3.x.x.tar.Z
   tar -xvf EMC.AIX.5.3.x.x.tar
   ```

9. Run the following command to install the XtremIO ODM filesets:

   `installp -acgXd . EMC.XtremIO.aix.rte EMC.XtremIO.fcp.MPIO.rte`

10. Run the following command to configure the hdisk devices:

    `cfgmgr`

Note: All devices are updated except `rootvg` devices. The `rootvg` devices are updated after the system reboot.

---

11. Run the following command to reboot the host so that the current `rootvg` hdisk devices are reconfigured as XtremIO devices:

    shutdown -Fr

12. Run the following command to ensure that the hdisks are configured as XtremIO devices:

    lsdev -Cc disk

    A message similar to the following example displays for each hdisk device that was successfully configured:

    ```
    hdisk2  Available 14-T1-01 EMC XtremIO FCP MPIO Disk
    hdisk3  Available 14-T1-01 EMC XtremIO FCP MPIO Disk
    hdisk4  Available 14-T1-01 EMC XtremIO FCP MPIO Disk
    hdisk5  Available 14-T1-01 EMC XtremIO FCP MPIO Disk
    ```

    The migration process is complete and you can start the applications.

Migration steps when booting from XtremIO PowerPath hdiskpower devices

1. Run the following command to turn off multipathing to the root device:

   pprootdev off

2. Run the following command to reboot the host so that the boot device is no longer under PowerPath control:

   shutdown -Fr

3. Remove any ghost disk devices that were created. Clean PowerPath first by running the **powermt check** command and when you see the following message, enter ***y*** (`yes`) to confirm the removal of any dead paths.

   ```
   Warning: XtremIO device path hdisk4 is currently dead.
   Do you want to remove it (y/n/a/q)? y
   ```

4. Run the following command to list and remove any XtremIO hdisk devices that are in the **Defined** state:

   lsdev -CtXtremIO -SD -F name | xargs -n1 rmdev -dl

5. Run the following command to remove all configured hdiskpower devices:

   ```
   powermt remove dev=all
   lsdev -Cc disk -F name | grep hdiskpower | xargs -n1 rmdev -dl
   ```

6. Complete the following steps to download the `boot_change_v4.tar` package from the ftp server:

   a. Run the following command to go to the ftp server:

      ftp ftp.emc.com

   b. Log in with an anonymous username and your e-mail address as password.

   c. Run the following command to go to the package location:

      cd /pub/elab/aix/ODM_DEFINITIONS

   d. Run the following command to download the package:

      get boot_change_v4.tar

e. Run the following command to exit:

```
quit
```

7. Run the following command to extract the files from the `boot_change_v4.tar` package:

```
tar –xvf boot_change_v4.tar
```

The following files will be extracted into the existing directory:

```
boot_change.README
boot_change_v4.sh
osfcpdisk.add
osfcpmpiodisk.add
osscsidisk.add
```

8. Run the following command to convert the existing XtremIO hdisk ODM definitions to **Other FC SCSI Disk Drive**:

```
./boot_change_v4.sh
```

The script will request input from the user. Enter **X** for XtremIO devices.

A message similar to the following example displays for each hdisk device that was successfully converted:

```
XTREMIO FCP device hdisk2 was converted to Other FC SCSI Disk Drive
XTREMIO FCP device hdisk3 was converted to Other FC SCSI Disk Drive
XTREMIO FCP device hdisk4 was converted to Other FC SCSI Disk Drive
XTREMIO FCP device hdisk5 was converted to Other FC SCSI Disk Drive
```

9. Run the following command to remove all converted hdisks and SACD hdisks that are not part of `rootvg`:

```
lsdev -Ct osdisk -F name | xargs -n1 rmdev -dl
lsdev -CtXIO* -F name | xargs -n1 rmdev -dl
```

A message similar to the following example displays for each hdisk device that was successfully removed:

```
hdisk2 deleted
hdisk3 deleted
hdisk4 deleted
hdisk5 deleted
```

10. Run the following command to uninstall the XtremIO ODM filesets:

```
installp -u EMC.XtremIO.aix.rte EMC.XtremIO.fcp.rte
```

11. Complete the following steps to download the AIX ODM Package Version 5.3.x.x or 6.x.x.x from the ftp server:

a. Run the following command to go to the ftp server:

```
ftp ftp.emc.com
```

b. Log in with an anonymous username and your e-mail address as password.

c. Run the following command to go to the AIX ODM package location:

```
cd /pub/elab/aix/ODM_DEFINITIONS
```

d. Run the following command to download the AIX ODM package:

```
get EMC.AIX.5.3.x.x.tar.Z
```

e. Run the following command to exit:

```
quit
```

12. Place the AIX ODM Package version 5.3.x.x in the **cd /usr/sys/inst.images** directory for installation.

13. Run the following commands to uncompress and untar the package:

```
uncompress EMC.AIX.5.3.x.x.tar.Z
tar -xvf EMC.AIX.5.3.x.x.tar
```

14. Run the following command to install the XtremIO ODM filesets:

```
installp -acgXd . EMC.XtremIO.aix.rte EMC.XtremIO.fcp.rte
```

15. Run the following command to configure the hdisk devices as XtremIO:

```
cfgmgr
```

Note: All the devices are updated except `rootvg` devices. The `rootvg` devices are updated after the system reboot.

16. Run the following command to reboot the host to configure the current hdisk `rootvg` devices as XtremIO devices:

```
shutdown -Fr
```

17. Run the following command to configure PowerPath devices:

```
powermt config
```

18. Run the following command to enable PowerPath as the boot device:

```
pprootdev on
```

19. Run the following command to reboot the host:

```
shutdown -Fr
```

Note: PowerPath now controls the XtremIO boot device.

The migration process is complete and you can start the applications.

Migration when migrating client from AIX 6.1 to AIX 7.1

Prerequisites

These operations assume all SAN disks are disks served up by NPIV virtual HBAs. The vscsi disks will be served up by a VIO server, and the steps to migrate VIOS EMC ODM definitions are basically the same steps on the VIO server.

Migration steps

1. Unmount all file systems and varyoff all volume groups by typing the following:

   ```
   varyoffvg vg01
   ```

2. Remove all Dell EMC SAN hdisks (which can also be done with `rmdev -Rdl fcsX` where X is the fcs number) by typing the following:

   ```
   lsdev -Cc disk -F name | xargs -n1 rmdev -dl
   ```

3. Uninstall all EMC ODM definitions by typing the following:

   ```
   lslpp -L | grep EMC
   installp -u EMC.CLARiiON*
   installp -u EMC. Symmetrix*
   ```

4. Verify that all definitions have been uninstalled.
5. Migrate AIX from 6.1 to 7.1 referring to IBM guide
6. Download the AIX ODM Package Version 6.0.0.x from the ftp server ftp.emc.com.
   a. Type the following command:

      ```
      ftp ftp.emc.com
      ```

   b. Log in with a username of anonymous and your email address as a password.
   c. Type the following command:

      ```
      cd /pub/elab/aix/ODM_DEFINITIONS
      ```

   d. Type the following command:

      ```
      get EMC.AIX.5.3.0.X.tar.Z
      ```

   e. Type the following command:

      ```
      quit
      ```

7. Place the AIX ODM Package Version 6.0.0.x into the **/usr/sys/inst.images** directory for installation. Uncompress and untar the package. Type:

   ```
   cd /usr/sys/inst.images
   uncompress EMC.AIX.6.0.0.X.tar.Z
   tar -xvf EMC.AIX.6.0.0.X.tar
   ```

8. Install the Symmetrix ODM filesets by typing:

   ```
   installp -acgXd . EMC.Symmetrix.aix.rte EMC.Symmetrix.fcp.rte
   ```

9. Configure the hdisk devices as Symmetrix by typing the following command. All devices will be updated except rootvg. This will be updated after the system reboot is performed in step 10.:

   /usr/lpp/EMC/Symmetrix/bin/emc_cfgmgr

10. Reboot the host so that the current hdisk rootvg devices will be reconfigured as Symmetrix devices. Type:

    shutdown -Fr

At this point, the migration process is complete. The customer may start their applications.

CHAPTER 4

Symantec Veritas Storage Foundation and AIX

This chapter provides information regarding Symantec Veritas Storage Foundation and AIX with Symmetrix, VNX series, and CLARiiON systems.

Overview

Dell EMC supports Symantec Veritas Storage Foundation for AIX with both Symmetrix, VNX series, or CLARiiON systems. This chapter addresses some information to consider if you are planning to use Veritas Volume Manager (VxVM) in your environment.

IMPORTANT

Dell EMC strongly recommends upgrading existing AIX environments running VxVM 3.X or 4.X, to VxVM 5.0 or later.

Dell EMC does *not* support active coexistence between PowerPath and VxVM Dynamic Multipathing (DMP). In other words, the use of PowerPath to control one set of storage array devices, and DMP to control another set of storage array devices within the same AIX server is *not* supported. This holds true for both Symmetrix, VNX series, or CLARiiON systems. If PowerPath is installed, *all* storage array devices *must* be controlled by PowerPath.

AIX 6.1 requires a special build from Symantec called *5.0MP1 Update 1 for AIX 6.1*. For more information, contact Symantec.

Always consult the Dell EMC Simple Support Matrices for the latest supported versions of Veritas Storage Foundation for AIX.

VMAX or Symmetrix

This section contains information for VxVM DMP and VxVM and PowerPath on DMX.

VxVM DMP

VxVM DMP requires the **Common Serial Number** (C bit) bit enabled on each FA director used by DMP.

The **SPC-2 Compliance** (SPC-2) director bit is required for microcode 5773 or later, or wherever it may be defined in the Dell EMC Simple Support Matrices.

Some versions of Veritas Storage Foundation do *not* support the SPC-2 director bit. The SPC-2 director bit can be disabled on Symmetrix arrays running microcode 5773 if the installed version of VxVM does not support the SPC-2 bit. This would include VxVM 4.0.2.0 (MP2) and earlier.

Note: Dell EMC recommends upgrading Veritas Storage Foundation to 5.0 or later.

VxVM and PowerPath

With the release of Veritas Storage Foundation 4.0.1.0 (MP1), DMP automatically detects the presence of PowerPath on an AIX server. This feature allows DMP to concede all multipathing duties to PowerPath without user intervention.

VxVM DMP and PowerPath on VNX series or CLARiiON

This section contains information for VxVM DMP and VxVM and PowerPath on VNX series or CLARiiON.

VxVM DMP

Some versions of VxVM are *not* supported with VNX series or CLARiiON systems. The Dell EMC Simple Support Matrices contains the proper versions supported with VNX series or CLARiiON systems. Always consult the Dell EMC Simple Support Matrices for the latest supported versions.

Dell EMC supports `failovermode 1` for VxVM 5.0.1.3 (MP1 RP3) and later. Prior to this version of VxVM, only `failovermode 2` is supported.

Dell EMC supports `failovermode 4` (ALUA), starting with VxVM 5.0.3.2 and later.

With VxVM 5.0.1.5 and earlier, Dell EMC currently supports a maximum of one AIX initiator per CLARiiON Storage Processor, per AIX server, with DMP.

With VxVM 5.0.3.0 and later, more than one AIX initiator per CLARiiON Storage Processor is supported.

Support for VNX series and CLARiiON NDU operations with DMP are not subject to the stipulations defined in the "Non-disruptive upgrades in a VNX series environment with PowerPath" on page 271. Check the Dell EMC Simple Support Matrices for supported NDU configurations with VxVM and AIX.

VxVM and PowerPath

With the release of Veritas Storage Foundation 4.0.1.0 (MP1), VxVM automatically detects the presence of PowerPath on an AIX server. This feature allows DMP to concede all multipathing duties to PowerPath without user intervention.

Please see "NDU checklist" on page 271 for more information regarding proper settings for NDU support with PowerPath.

Symantec Veritas Storage Foundation feature functionality

This section contains general support rules and guidelines regarding special features and functionality available from the Veritas Storage Foundation product. Always consult the Dell EMC Simple Support Matrices for the latest supported versions of Veritas Storage Foundation for Microsoft Windows.

Review the *Veritas Storage Foundation Advanced Features Administrator's Guide* for more details.

Thin Reclamation (VxVM)

Array-based Thin Reclamation is supported with EMC VMAX 400K, VMAX 200K, VMAX 100K, VMAX 40K, VMAX 20K/VMAX, VMAX 10K (Systems with SN xxx987xxxx), VMAX 10K (Systems with SN xxx959xxxx), VMAXe, and CLARiiON systems and requires Storage Foundation 6.1 and later. VxVM Thin Reclamation functionality is supported with minimum FLARE Release 29, version 04.29.000.5.003 or later. Thin Reclamation functionality is performed on a per-LUN basis.

Host-based Thin Reclamation functionality is supported with VMAX 400K, VMAX 200K, VMAX 100K, VMAX 40K, VMAX 20K/VMAX, VMAX 10K (Systems with SN xxx987xxxx), VMAX 10K (Systems with SN xxx959xxxx), VMAXe, VNX series, and CLARiiON. For VMAX and VMAXe, host-based Thin Reclamation was introduced with Enginuity 5875.135.91.

For IBM operating systems AIX 6.1 and AIX 7.1, host-based Thin Reclamation requires the following software components:

- Symantec Veritas Storage foundation 6.1 SP1 and later
- DMP multipath software
- VMAX 400K with HYPERMAX 5977.250.189 or later
- VMAX 200K with HYPERMAX 5977.250.189 or later
- VMAX 100K with HYPERMAX 5977.250.189 or later
- VMAX 40K with Enginuity 5876.82.57 or later
- VMAX 20K with Enginuity 5876.82.57 or later
- VMAX with Enginuity 5875.135.91 or later
- VMAX 10K (Systems with SN xxx987xxxx) with Enginuity 5876.159.102 or later
- VMAX 10K (Systems with SN xxx959xxxx) with Enginuity 5876.159.102 or later
- VMAXe with Enginuity 5875.198.148e or later
- CX-4 with FLARE 04.30.000.5.509 or later
- VNX with VNX OE for Block v31 or later

Note: Refer to the Dell EMC Simple Support Matrices, available at Dell E-Lab Navigator, for the most up-to-date support information. EMC PowerPath is not supported with host-based Thin Reclamation at this time.

SmartMove (VxVM)

VxVM SmartMove is supported with VNX series systems, CLARiiON CX4 and newer arrays, and Symmetrix DMX-3, DMX-4, VMAX400K, VMAX 200K, VMAX100K, VMAX 40K, VMAX 20K/VMAX, VMAX 10K (Systems with SN xxx987xxxx), VMAX 10K (Systems with SN xxx959xxxx), and VMAXe arrays.

EMC supports SmartMove with minimum VxVM 5.1.

PART 1

Symmetrix Connectivity

Part 2 includes:

CHAPTER 5

AIX, VMAX/PowerMax Family Environment

This chapter provides information specific to IBM AIX hosts connecting to VMAX/PowerMax systems.

Note: Any general references to Symmetrix or Symmetrix array include the DMX, VMAX, VMAX2, VMAX3, VMAX All Flash, and PowerMax family.

AIX, VMAX/PowerMax Family Environment

This section provides an overview of hardware and software support for AIX hosts connecting to EMC VMAX/PowerMax systems.

Hardware connectivity

Refer to the Dell EMC Simple Support Matrices or contact your Dell EMC representative for the latest information on qualified hosts, host bus adapters, and connectivity equipment.

Clustering software

The following software is supported:

- HACMP (High-Availability Cluster Multi-Processing)
- RVSD (Recoverable Virtual Shared Disk)
- Spectrum Scale (Formerly GPFS)

Logical devices

LUNs supported are shown in Table 10.

Table 10 Symmetrix LUN support

Symmetrix model	Maximum LUNs
3330/5330 3430/5430 3630/5630 3700/5700 3830/5830 3930/5930	128
8230 8430/8530 8730/8830	256
DMX/DMX-2 Series	512
DMX-3 Series	2048
VMAX 40K, VMAX 20K/VMAX, VMAX 10K (Systems with SN xxx987xxxx), VMAX 10K (Systems with SN xxx959xxxx), and VMAXe.	EMC Knowledgebase solution emc67373

Making a volume group

Note: In the following examples, each Symmetrix disk device will be defined as a one-volume group for simplicity. You can choose to treat each Symmetrix device as one volume group or join multiple Symmetrix devices into one volume group.

To make a volume group for a Symmetrix device, enter a statement similar to the following for each volume group:

```
mkvg -y testvg hdisk2
```

For example, the above command creates a volume group from the physical disk device `hdisk2` named `testvg`.

The default physical partition size is 4 MB.

Making the volume group available

After a volume group is created, it must be made available to the operating system.

To make a volume group available to the host, use `SMIT` or type the following statement:

```
varyonvg <volume_group_name>
```

Repeat this action for each volume group. The **varyonvg** command activates a volume group, making it known to the host system. (Conversely, the **varyoffvg** command deactivates a volume group, making it unavailable to the host.)

Creating logical volumes

Now that the volume groups are defined, you can either use the volume groups as is or define logical volumes.

The following example illustrates how to make the logical volume `testlv` in volume group `testvg` as a journaled file system with 100 physical partitions, 400 MB in size.

```
mklv -y testlv -t jfs testvg 100
```

Creating a mount point

Create a mount point that points to a logical volume. The following command creates a mount point named `emc`:

```
mkdir /emc
```

Creating and mounting a JFS file system

To create and mount file systems use `SMIT` or the `crfs` and `mount` commands.

For example, use the **crfs** command in a statement similar to the following to create a file system:

```
crfs -v jfs -d /dev/testlv -m /emc -A yes -a size=41943094
```

where:

-v jfs	Makes a journaled file system.
-d	Specifies the logical volume.
-m /emc	Specifies the mount point for the file system; in this case `/emc`.
-A yes	Mounts file system automatically at system restart.
-a	Specifies the various attributes for this file system.
size=x	Specifies the file system size in blocks (512 bytes per block). 41943094 is the size of a 2 GB file system.

- Check the file system as follows:

  ```
  fsck /emc
  ```

  The **fsck** command checks the integrity of the file system mounted at the `/emc` directory.

- Mount each file system at the desired mount point as follows:

  ```
  mount /emc
  ```

- Change to the `/emc` directory, and then create a directory called `lost+found`:

  ```
  cd /emc
  mk lost+found
  ```

Configuring disks using SMIT

Previous sections of this chapter described adding devices to the AIX server using standalone commands. You can also add devices using the **SMIT** menu.

To access the **SMIT** menu:

1. Log in as `root`.
2. At the system prompt, type **SMIT** and press **Enter**.

Adding a volume group

Follow these steps to create and name a volume group and to select the physical volume(s) for that group:

1. Select **System Storage Management** from the **SMIT** main menu.
2. Select **Logical Volume Manager** from the menu.
3. Select **Volume Groups** from the menu.
4. Select **Add a Volume Group** from the menu.
5. Enter a volume group name (**`vg01`**, for example) when prompted.
6. Tab down to the next field and enter the physical partition size for the disk. (The default size is 4 MB.)
7. Tab down, and press **F4**. Select the physical volume name by pressing **F7** when you are next to the disk(s) you want to select; then press **Enter**.
8. Press **Enter** again. If this action completes successfully, press **F3** three times to return to the **Logical Volume Manager** menu.

Adding logical volumes to the volume group

Follow these steps to add logical volumes to the volume group:

1. Select **Logical Volumes** from the **Logical Volume Manager** menu.
2. Select **Add a Logical Volume**.
3. Press **F4** and select the volume group to which you are adding the logical volume(s) by pressing **F7** when beside the volume group you wish to select. Then press **ENTER**.

   Note: If you are adding the logical volume to a new volume group, simply enter a name for the volume group.

4. Enter a name for the logical volume when prompted.
5. Enter the number of physical partitions.
6. Enter the physical volume name. Pressing **F4** displays a pop-up list of available volume groups.
7. Change any other parameters as desired, such as striping options.
8. Press **Enter**. If this action completes successfully, press **F3** four times to return to the **Physical and Logical Storage** menu.

Creating file systems

Perform the following steps to create and mount file systems:

1. Select **File Systems** from the **Physical and Logical Storage** menu.
2. Select **Add/Change/Show/Delete File Systems** from the menu.
3. Select **Journaled File System** (jfs) to create a journaled file system.
4. Select **Add a Journaled File System on a Previously Defined Logical Volume** from the menu that appears.
5. Select **Add a Standard Journaled File System** from the menu.
6. Enter the name of logical volume on which to create a file system. Pressing **F4** displays a pop-up list of available logical volumes.
7. Enter a mount point for that file system.
8. Click **Enter**. If this action completes successfully, press **Cancel** if you are using an X terminal (or **F3** several times if using an ASCII terminal) to return to the **File System** menu.
9. Select **Mount a File System** from the menu.
10. Select the file system you wish to mount. Pressing **F4** displays a pop-up list of available file systems.
11. Select **Mount Point** from the menu that appears. Pressing **F4** displays a pop-up list of available mount points.
12. Click **DO**. If this action completes successfully, press **Cancel** if you are using an X terminal (or **F3** if using an ASCII terminal) to exit `SMIT`.
13. Verify the file system is mounted by issuing the **`mount`** or **`df`** command.

Creating and mounting an enhanced JFS2 file system

To create and mount file systems use `SMIT` or the **`crfs`** and **`mount`** commands. For example, use the **`crfs`** command in a statement similar to the following to create a file system:

```
crfs -v jfs2 -d /dev/testlv -m /emc -A yes -p rw -a agblkssize=4096
```

where:

`-v jfs2`	Makes an enhanced journaled file system.
`-d`	Specifies the logical volume.
`-m /emc`	Specifies the mount point for the file system; in this case `/emc`.
`-A yes`	Mounts file system automatically at system restart.
`-p`	Sets the permissions for this file system.
`-a`	Specifies the various attributes for this file system.
`agblkssize`	Specifies the JFS2 block size in bytes.

- Check the file system as follows:

  ```
  fsck /emc
  ```

  The **`fsck`** command checks the integrity of the file system mounted at the `/emc` directory.

- Mount each file system at the desired mount point as follows:

  ```
  mount /emc
  ```

- Change to the `/emc` directory, and then create a directory called `lost+found`:

  ```
  cd /emc
  mk lost+found
  ```

Creating file systems

Perform the following steps to create and mount file systems:

1. Select **File Systems** from the **Physical and Logical Storage** menu.
2. Select **Add/Change/Show/Delete File Systems** from the menu.
3. Select **Enhanced Journaled File System** (jfs) to create a journaled file system.
4. Select **Add an Enhanced Journaled File System on a Previously Defined Logical Volume** from the menu that appears.
5. Select **Add an Enhanced Journaled File System** from the menu.
6. Enter the name of logical volume on which to create a file system. Pressing **F4** displays a pop-up list of available logical volumes.
7. Enter a mount point for that file system.
8. Click **Enter**. If this action completes successfully, press **Cancel** if you are using an X terminal (or **F3** several times if using an ASCII terminal) to return to the **File System** menu.
9. Select **Mount a File System** from the menu.
10. Select the file system you wish to mount. Pressing **F4** displays a pop-up list of available file systems.
11. Select **Mount Point** from the menu that appears. Pressing **F4** displays a pop-up list of available mount points.
12. Click **DO**. If this action completes successfully, press **Cancel** if you are using an X terminal (or **F3** if using an ASCII terminal) to exit `SMIT`.
13. Verify the file system is mounted by issuing the **`mount`** or **`df`** command.

Adding devices online

Whenever devices are added online to the Symmetrix system or device channel addresses are changed, you must run the `emc_cfgmgr` script to introduce the new devices to the system:

1. Change directory to `/usr/lpp/EMC/Symmetrix/bin`.
2. Invoke the shell script `emc_cfgmgr`.

Whenever TimeFinder devices are configured by the host, they are forced to a defined state. To configure all defined Symmetrix TimeFinder devices and set them to an available state, enter:

```
mkbcv -a
```

LUN expansion

For existing AIX file systems to recognize and take advantage of any additional capacity provided by Symmetrix LUN expansion, AIX servers must have AIX Revision 5.2 ML-01 APAR IY21957 or later installed.

The steps for performing this expansion are:

1. Run the **inq** command and note the initial size of the device.
2. Perform LUN expansion (RAID Group or MetaLUN) as detailed in the appropriate Symmetrix xxxx product guide for the Symmetrix model.
3. From AIX, unmount the file system and vary off the volume group (using the **umount** and **varyoffvg** commands).
4. From AIX, run the **emc_cfgmgr** command.
5. Run the **inq** command to verify that AIX recognizes the hdisk's increased capacity.
6. Vary on the volumegroup with the **varyonvg** command.
7. Run the **chvg -g** <*vg_name*> command so the volume group's new capacity is recognized.
8. Run the **lslv** <*lv_name*> command and note the PP SIZE and PPs. Multiply the two in order to approximate the new size of the volume in megabytes. Convert this number to 512-byte blocks by multiplying by 2048. (If your logical volume PP SIZE is less than 1 MB, then adjust the calculation accordingly.)
9. Mount the file system, and run the **chfs -a size=**<*new_size_from_step_8*> <*fs_name*> command. This will expand the file system, using all new space available on the device.

System and error messages

In an AIX environment, you can view system status and error messages by using either `SMIT` or the standalone command **`errpt`.**

AIX logs system status and error messages in `/var/adm/ras/errlog`. This error log is not viewable directly.

Viewing messages using SMIT

Follow these steps to view the error log using `SMIT`:

1. Select **Problem Determination** from the `SMIT` main menu.
2. Select **Error Log**.
3. Select **Generate Error Report**.
4. Decide where the generated report will reside.
5. Respond **Yes** or **No** at the **Concurrent Error Reporting** prompt.
6. Modify the reporting options as appropriate for your environment and press **Enter** to generate the report.

Viewing messages using errpt

The **`errpt`** command has the following format:

`errpt` **`-`***`<options> <parameters>`*

Table 11 on page 204 lists the `errpt` options and parameters.

Table 11 Options and parameters for the errpt command

Command options and parameters	Definition
`-a`	Displays information about errors in the error log file in a detailed format.
`-c`	Formats and displays each of the error entries concurrently.
`-e` ***`Enddate`***	Specifies all records posted before the *Enddate* variable, which has the format `mmddhhmmyy` (month, day, hour, minute, and year).
`-s` ***`Startdate`***	Specifies all records posted after the *Startdate* variable, which has the format `mmddhhmmyy` (month, day, hour, minute, and year).
`-S` ***`ResourceClassList`***	Generates a report of resource classes specified by the *ResourceClassList* variable. Some *ResourceClassList* members are disk, adapter, and tape.
`-j`	Displays information about a specific error ID.
`None`	Displays a complete summary report.

Example The following is an example of an `errpt` output. Note that error report entries vary considerably, and yours might not resemble this example:

```
LABEL:SC_DISK_ERR3
IDENTIFIER:C62E1EB7

Date/Time:Mon Aug 25 11:57:31 EDT
Sequence Number:325591
Machine Id:000DBA1D4C00
Node Id:          182bc115      <- Host on which the error occurred
Class:            H             <- Hardware-classified error
Type:             PERM          <- Non-recoverable error
Resource Name:    hdisk29       <- Disk that reported the error
Resource Class:   disk          <- Resource that identified the error
Resource Type:    SYMM_RAID1    <- Resource type defined in the CuDv ODM Object Class
Location:         P1-I5/Q1-W5006048ACADFE223-L43C000000000000 <- World Wide Name
VPD:
     Manufacturer.................EMC
     Machine Type and Model......SYMMETRIX
     ROS Level and ID............5669          <- Symmetrix Enginuity version
     Serial Number...............7244A361      <- Symmetrix device serial number
     Part Number.................00000000000610023010187
     EC Level....................490072        <- Last six bytes of Symmetrix unit
     Device Specific.(Z0).........00
     Device Specific.(Z1).........61
     Device Specific.(Z2).........566900460000000000060503 <- Enginuity version/date
     Device Specific.(Z3).........12000000      <- Device-specific data
     Device Specific.(Z4).........54110008      <- Device-specific data
     Device Specific.(Z5).........4F80          <- Symmetrix unit cache size in hex

Description
DISK OPERATION ERROR

Probable Causes
DASD DEVICE
STORAGE DEVICE CABLE

Failure Causes
DISK DRIVE
DISK DRIVE ELECTRONICS
STORAGE DEVICE CABLE

       Recommended Actions
       PERFORM PROBLEM DETERMINATION PROCEDURES

Detail Data
PATH ID
SENSE DATA
0600 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0200 0400 0000 0000
0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000
0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000
0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000
0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000
0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000
0000 0000 0000 0000 7E5F 0006 8C40 0000 00A3 0000 0000 0000 0000 0000 0000 0000
0000 0000 0000
```

FC SCSI-3 disk error log detail data

As shown in the example, the format of the sense data is:

```
SENSE DATA
LL00 CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC RRRR RRRR RRRR VVSS AARR DDDD KKDD
DDDD DDDD DDDD DDDD PPQQ DDDD DDDD DDDD DDDD DDDD DDDD DDDD DDDD DDDD DDDD DDDD
DDDD DDDD DDDD DDDD DDDD DDDD DDDD DDDD DDDD DDDD DDDD DDDD DDDD DDDD DDDD DDDD
DDDD DDDD DDDD DDDD DDDD DDDD DDDD DDDD DDDD DDDD DDDD DDDD DDDD DDDD DDDD DDDD
DDDD DDDD DDDD DDDD DDDD DDDD DDDD DDDD DDDD DDDD DDDD DDDD DDDD DDDD DDDD DDDD
DDDD DDDD DDDD DDDD DDDD DDDD DDDD DDDD DDDD DDDD DDDD DDDD DDDD DDDD DDDD DDDD
DDDD DDDD BBBB BBBB BBBB BBBB NNNN NNNN XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX
XXXX XXXX
```

The fields can be interpreted as follows:

Field	Description
LL	Length of the SCSI command that failed in the CC fields.
C	Bytes of the SCSI command that failed. The first two digits of the this field indicate the SCSI opcode.
VV	Status validity field, with these possible values: **00** Indicates no status validity was set by the adapter driver. **01** Indicates the SCSI status field (SS) is valid **02** Indicates the adapter status field (AA) is valid. **03** Indicates driver status is valid. This is used for when the device driver detects special errors not directly related with hardware errors.
SS	SCSI status field: **02** Check condition; indicates that the device had an error; addition information should be included in the detail sense fields marked with **D**, **P**, and **Q**. **08** The SCSI device is busy for some reason. This sometimes happens when devices are being shared between hosts and one of them is involved in error recovery. **18** Indicates that the SCSI device is reserved by another host.

Field	Description
AA	Adapter status field: **01** Host I/O bus error; indicates a hardware problem with either the SCSI adapter or the Microchannel/PCI bus. **02** Transport fault; indicates a hardware problem, which is probably due to problems with the SCSI transport layer (connection cables). **03** Command timeout; indicates that the SCSI command did not finish within the time it was supposed to. This usually indicates a hardware problem that is due to SCSI transport layer problems. **04** No device response; indicates that the SCSI devices did not reply to the issued SCSI command. This is a hardware problem that can be due to any of several problems, such as: bad SCSI device, power supply problem for the SCSI device, SCSI cabling and/or termination problems. **07** World Wide Name change; indicates that the SCSI devices at this ID has a different WWN and, therefore, is a different device.
R	Reserved for future use. Fields marked with an **R** should be **0**.
D	Part of the SCSI request sense data. Fields marked with a **D** are valid only when the SCSI status field (SS) has a value of **02**.
KK	Sense Key, which is part of the SCSI request sense data. Some common values are: **01** The device is indicating it was able to do the I/O after some error recovery. **02** The device is not ready. **03** Media error. **04** Hardware error at the device.
PP	Additional Sense Code (ASC), which is part of the SCSI request sense data.
QQ	Additional Sense Code Qualifier (ASCQ), which is part of the SCSI request sense data.
N	32-bit open count field (the number times the device has been opened when it was not already open).

CHAPTER 6

AIX, VMAX/PowerMax Family over Fibre Channel

This chapter provides information specific to AIX hosts connecting to VMAX/PowerMax systems over Fibre Channel.

Note: Any general references to Symmetrix or Symmetrix array include the DMX, VMAX, VMAX2, VMAX3, VMAX All Flash, and PowerMax family.

AIX, VMAX/PowerMax Family Fibre Channel environment

This section provides an overview of hardware and software support for AIX hosts connecting to Symmetrix systems over Fibre Channel.

Refer to Chapter 5, "AIX, VMAX/PowerMax Family Environment," for information that covers both the Fibre Channel and SCSI environments.

Boot device support

AIX hosts have been qualified for booting from Symmetrix devices interfaced over Fibre Channel as described in the Dell EMC Simple Support Matrices.

Upgrading Fibre Channel microcode

For more information about upgrading your host's Fibre Channel microcode in the host, refer to:

http://www14.software.ibm.com/webapp/set2/firmware/gjsn

Host configuration

In any AIX/VMAX/Symmetrix Fibre Channel environment using the IBM Native Fibre Channel driver with supported IBM host bus adapters (HBAs), follow the procedures in this section.

The necessary drivers for the adapter are included in the AIX operating system. Refer to the Dell EMC Simple Support Matrices for the latest PTF and firmware revisions.

Note: Refer to your AIX host documentation for specifications relating to PCI bus limitations and adapter location.

The following filesets are distributed within the ODM support package. Always refer to the Dell EMC Simple Support Matrices for the most up-to-date supported solutions:

- Symmetrix AIX Support Software
  - Standard utility support
- Symmetrix FCP Support Software
  - IBM Fibre Channel driver support.
  - Supports PowerPath
- Symmetrix FCP MPIO Support Software
  - IBM default PCM MPIO support
- Symmetrix iSCSI Support Software
  - IBM iSCSI driver support
  - Supports PowerPath
- Symmetrix HA Concurrent Support
  - Requires HACMP CLVM
- CLARiiON AIX Support Software
  - Standard utility support for VNX series and CLARiiON
- CLARiiON FCP Support Software
  - IBM Fibre Channel driver support for VNX series and CLARiiON
  - Supports PowerPath for VNX series and CLARiiON
- CLARiiON FCP MPIO Support Software
  - IBM default PCM MPIO support for VNX series and CLARiiON
- CLARiiON iSCSI Support Software
  - IBM iSCSI driver support for VNX series and CLARiiON
  - Supports PowerPath for VNX series and CLARiiON
- CLARiiON HA Concurrent Support
  - Requires HACMP CLVM for VNX series and CLARiiON

- Invista AIX Support Software
  - Standard utility support
- Invista FCP Support Software
  - IBM Fibre Channel driver support.
  - Supports PowerPath
- Celerra AIX Support Software
  - Standard utility support
- Celerra FCP Support Software
  - IBM iSCSI driver support

Installing the ODM support package

Perform the following procedure if you are installing the IBM Fibre Channel Interface Support for Symmetrix ODM Support Package.

1. Log on to the AIX server as **root**.
2. Obtain the ODM package from the FTP server:
   a. **ftp ftp.emc.com**
   b. **cd /pub/elab/aix/ODM_DEFINITIONS**
   c. **bin**
   d. **get EMC.AIX.5.2.X.tar.Z** (AIX 5.2)
   or
   get EMC.AIX.5.3.X.tar.Z (AIX 5.3 or 6.1)

   Note: These are the latest filesets as of this printing. If a more recent fileset is available, use it.

   e. **bye**
3. In the /tmp directory, uncompress the tar package:

   uncompress EMC.AIX.*X.X.X.X*.tar.Z

   where *X.X.X.X* is the fileset version, as noted in step 2d.
4. Extract the tar package:

   tar -xvf EMC.AIX.*X.X.X.X*.tar

   where *X.X.X.X* is the fileset version.
5. Step 5 depends on whether you want to complete the installation using SMIT or the command line. Proceed to:
   - "Using SMIT", next, or
   - "Using the command line" on page 214

Using SMIT

After performing Step 1 through Step 4 outlined in "Installing the ODM support package" beginning on page 213, complete the installation using SMIT:

5. At the command line, invoke SMIT from the /tmp directory:

   smitty installp
6. Select **Install and Update from Latest Available Software**.
7. Select the input device to be the current working directory:

   directory
   ./
8. Select **F4=List** to list available software.

9. Select **EMC Symmetrix AIX Support Software** using **F7=Select**.
   - For a Fibre Channel environment, also select **EMC Symmetrix Fibre Channel Support Software**.

     or

   - For a Fibre Channel MPIO environment, also select **EMC Symmetrix Fibre Channel MPIO Support Software**.

     Note: MPIO is supported only on AIX 5.2.00-02 or higher.

10. Select **Enter** to perform the action.
11. Accept the default options and press **Enter** to initiate the installation.
12. Scroll to the bottom of the installation summary screen to verify that the `SUCCESS` message is displayed.
13. Exit `SMIT`.

Using the command line

After performing Step 1 through Step 4 outlined in "Installing the ODM support package" beginning on page 213, complete the installation from the command line interface:

5. From the `/tmp` directory, enter the following command:

   `installp -ac -gX -d . EMC.Symmetrix.aix.rte EMC.Symmetrix.fcp.rte`

   where:

   - The period between `-d` and `EMC.Symmetrix` is the current directory where the installation file resides.
   - **`EMC.Symmetrix.aix.rte`** and **`EMC.Symmetrix.fcp.rte`**

     or

   - **`EMC.Symmetrix.aix.rte`** and **`EMC.Symmetrix.fcp.MPIO.rte`** are the names of the software to install as determined by the selections in Step 9.

6. Verify that the installation summary reports `SUCCESS`.

What next? Connect the Symmetrix system to the server, and proceed to "Configuring new devices" on page 214.

Configuring new devices

After you install the ODM software, you can run `emc_cfgmgr` to configure the devices:

1. Change (`cd`) to the directory `/usr/lpp/EMC/Symmetrix/bin`.
2. Invoke `emc_cfgmgr`.
3. If EMC TimeFinder is also installed, invoke the script `mkbcv -a`.

   Note: mkbcv is not required with VMAX 400K, VMAX 200K, VMAX 100K, VMAX 40K, VMAX 20K/VMAX, VMAX 10K (Systems with SN xxx987xxxx), VMAX 10K (Systems with SN xxx959xxxx), or VMAXe VRAID devices.

Addressing VMAX or PowerMax devices

This section describes the methods of addressing VMAX or PowerMax devices.

Arbitrated loop addressing

The Fibre Channel arbitrated loop (FC-AL) topology defines a method of addressing ports, arbitrating for use of the loop, and establishing a connection between Fibre Channel NL_Ports (level FC-2) on HBAs in the host and Fibre Channel directors (via their adapter cards) in the Symmetrix system. Once loop communications are established between the two NL_Ports, device addressing proceeds in accordance with the SCSI-3 Fibre Channel Protocol (SCSI-3 FCP, level FC-4).

The Loop Initialization Process (LIP) assigns a physical address (AL_PA) to each NL_Port in the loop. Ports that have a previously acquired AL_PA are allowed to keep it. If the address is not available, another address may be assigned, or the port may be set to non-participating mode.

Note: The AL-PA is the low-order 8 bits of the 24-bit address. (The upper 16 bits are used for Fibre Channel fabric addressing only; in FC-AL addresses, these bits are **x'0000'**.)

Symmetrix Fibre Channel director parameter settings control how the Symmetrix system responds to the Loop Initialization Process.

After the loop initialization is complete, the Symmetrix port can participate in a logical connection using the hard-assigned or soft-assigned address as its unique AL_PA. If the Symmetrix port is in non-participating mode, it is effectively off line and cannot make a logical connection with any other port.

A host initiating I/O with the Symmetrix system uses the AL_PA to request an open loop between itself and the Symmetrix port. Once the arbitration process has established a logical connection between the Symmetrix system and the host, addressing specific logical devices is done through the SCSI-3 FCP.

FC-AL addressing parameters

Refer to "Fibre Bit Settings" in the Dell EMC Simple Support Matrices for Symmetrix director flag settings.

Fabric addressing

Each port on a device attached to a fabric is assigned a unique 64-bit identifier called a World Wide Port Name (WWPN). These names are factory-set on the HBAs in the hosts, and are generated on the Fibre Channel directors in the Symmetrix system.

Note: For comparison to Ethernet terminology, an HBA is analogous to a NIC card, and a WWPN to a MAC address.

Note: The ANSI standard also defines a World Wide Node Name (WWNN), but this name has not been consistently defined by the industry.

When an N_Port (host server or storage device) connects to the fabric, a login process occurs between the N_Port and the F_Port on the fabric switch. During this process, the devices agree on such operating parameters as class of service, flow control rules, and fabric addressing. The N_Port's fabric address is assigned by the switch and sent to the N_Port. This value becomes the Source ID (SID) on the N_Port outbound frames and the Destination ID (DID) on the N_Port inbound frames.

The physical address is a pair of numbers that identify the switch and port, in the format **s,p**, where **s** is a domain ID and **p** is a value associated to a physical port in the domain. The physical address of the N_Port can change when a link is moved from one switch port to another switch port. The WWPN of the N_Port, however, does not change. A Name Server in the switch maintains a table of all logged-in devices, so N_Ports can automatically adjust to changes in the fabric address by keying off the WWPN.

The highest level of login that occurs is the process login. This is used to establish connectivity between the upper-level protocols on the nodes. An example is the login process that occurs at the SCSI FCP level between the HBA and the Symmetrix system.

SCSI-3 FCP addressing

The Symmetrix director extracts the SCSI Command Descriptor Blocks (CDB) from the frames received through the Fibre Channel link. Standard SCSI-3 protocol is used to determine the addressing mode and to address specific devices.

The Symmetrix system supports three addressing methods based on a single-layer hierarchy as defined by the SCSI-3 Controller Commands (SCC):

- Peripheral Device Addressing
- Logical Unit Addressing
- Volume Set Addressing

All three methods use the first two bytes (0 and 1) of the 8-byte LUN addressing structure. The remaining six bytes are set to 0s.

For Logical Unit and Volume Set addressing, the Symmetrix port identifies itself as an Array Controller in response to a host's Inquiry command sent to LUN 00. This identification is done by returning the byte 0x0C in the **Peripheral Device Type** field of the returned data for Inquiry. If the Symmetrix system returns the byte 0x00 in the first byte of the returned data for Inquiry, the Symmetrix system is identified as a Direct Access device.

Upon identifying the Symmetrix system as an Array Controller device, the host should issue a SCSI-3 **Report LUNS** command (0xA0), in order to discover the LUNS.

The three addressing modes, contrasted under "Symmetrix SCSI-3 addressing modes" on page 217, differ in the addressing scheme (target ID, LUN and virtual bus) and number of addressable devices.

Symmetrix SCSI-3 addressing modes

Table 12 lists the Symmetrix SCSI-3 addressing modes.

Table 12 Symmetrix SCSI-3 addressing modes

Addressing mode	Code[1]	"A" Bit	"V" Bit	Response to "Inquiry"	LUN Discovery method	Possible addresses	Maximum logical devices[2]
Peripheral Device	00	0	X	0x00 Direct Access	Access LUNs directly	16,384	256
Logical Unit	10	1	0	0x0C Array Controller	Host issues "Report LUNS" command	2,048	128
Volume Set	01	1	1	0x0C Array Controller	Host issues "Report LUNS" command	16,384	512

1. Bits 7-6 of byte 0 of the address.
2. The actual number of supported devices may be limited by the type host or host bus adapter used.

Note: The addressing modes are provided to allow flexibility in interfacing with various hosts. In all three cases the received address is converted to the internal Symmetrix addressing structure. Volume Set addressing is the default for Symmetrix systems. Select the addressing mode that is appropriate to your host.

Symmetrix SPC-2 considerations

This section contains information to consider and support for the SPC-2 director bit.

SPC-2 director bit considerations

The latest releases of Enginuity code 5671 for DMX and DMX-2, and Enginuity code 5771 for DMX-3, provide support for compliance with newer SCSI protocol specifications, specifically SCSI Primary Commands - 2 (SPC-2) as defined in the SCSI document located at http://www.t10.org.

The SPC-2 implementation in Enginuity includes functionality which, based on OS and application support, may enhance disk attach behavior to use newer SCSI commands which are optimized for a SAN environment (as implemented in Fibre Channel), as opposed to legacy (non SPC-2) functionality, which was targeted for older SCSI implementations utilizing physical SCSI bus-based connectivity (which cannot leverage the enhanced functionality of newer SCSI specifications).

SPC-2 director bit and AIX

At this time, the SPC-2 bit is not a requirement for AIX connectivity. Enabling the SPC-2 bit on an AIX 5L attached front-end director port has no impact to AIX connectivity in general and is allowed, for example, on director ports shared across other operating system platforms that require this setting.

IMPORTANT

Dell EMC highly recommends enabling the SPC-2 director bit for all new and existing AIX Fibre Channel and iSCSI attached configurations.

The one exception for AIX and the SPC-2 bit is with servers running Veritas Volume Manager. For SPC-2 support with VxVM, minimum VxVM 4.0 MP4 is required. The SPC-2 bit allows the Veritas Volume Manager to discover Symmetrix LUNs addressed above 8192.

Note: Refer to EMC Knowledgebase solution emc179861 for help enabling SPC-2 functionality in existing AIX environments running Veritas Volume Manager.

Always check the Dell EMC Simple Support Matrices for the most up-to-date support information.

AIX Multiple Path I/O

Multiple Path I/O (MPIO) allows for a device to be detected uniquely through one or more physical connections, or paths. A path-control module (PCM) provides the path management functions.

An MPIO-capable device driver can control more than one type of target device. A PCM can support one or more specific devices. Therefore, one device driver can be interfaced to multiple PCMs that control the I/O across the paths to each of the target devices.

Figure 9 shows the interaction among the different components that make up the MPIO solution. In this figure, the MPIO device driver controls multiple types of target devices, each requiring a different PCM. (KE=Kernel Extension, RTL=Run-time Loadable)

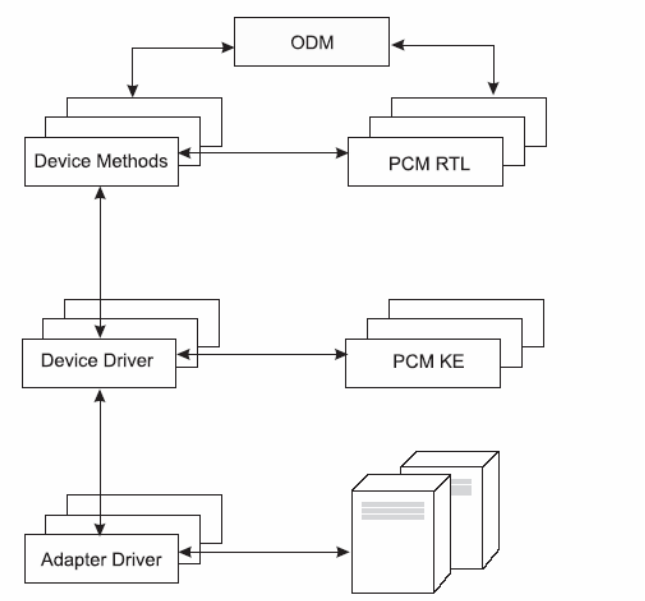


Figure 9 The MPIO solution

Before a device can take advantage of MPIO, the device's driver, methods, and predefined attributes in the Object Database Manager (ODM) must be modified to support detection, configuration, and management of multiple paths. Beginning with AIX 5.2 with the 5200-02 Recommended Maintenance package, the AIX Fibre Channel disk device drivers and their device methods support MPIO disk devices.

The AIX PCM consists of: the PCM RTL configuration module and the PCM KE kernel extension. The PCM RTL is a run-time loadable module that enables the device methods to detect additional PCM KE device-specific or path ODM attributes that the PCM KE requires. The PCM RTL is loaded by a device method. One or more routines within the PCM RTL are then accessed to perform specific operations that initialize or modify PM KE variables.

The PCM KE supplies path-control management capabilities to any device driver that supports the MPIO interface. The PCM KE depends on device configuration to detect paths and communicate that information to the device driver. Each MPIO-capable device driver adds the paths to a device from its immediate parent or parents. The maintenance and scheduling of I/O across different paths is provided by the PCM KE and is not apparent to the MPIO-capable device driver.

The PCM KE can provide more than one routing algorithm, which can be selected by the user. The PCM KE also helps collect information that can be used to determine and select the best path for any I/O request. The PCM KE can select the best path based on a variety of criteria, including load balancing, connection speed, connection failure, and so forth.

The AIX PCM has a health-check capability that can be used to do the following: Check the paths and determine which paths are currently usable for sending I/O, Enable a path that was previously marked failed because of a temporary path fault (for example, when a cable to a device was removed and then reconnected), Check currently unused paths that would be used if a failover occurred (for example, when the algorithm attribute value is failover, the health check can test the alternate paths).

Configuring an MPIO device

Configuring an MPIO-capable device uses the same commands as a non-MPIO device. At AIX 5.2 with 5200-02, the **`cfgmgr`**, **`mkdev`**, **`chdev`**, **`rmdev`**, and **`lsdev`** commands support managing MPIO device instances and display their attributes. An MPIO device instance also has path instances associated with the device instance. The **`mkpath, chpath, rmpath`**, and **`lspath`** commands manage path instances and display their attributes. A path instance can be added or removed from a MPIO device without unconfiguring the device.

An MPIO device can have additional attributes, but the required attributes that all MPIO devices must support are `reserve_policy` and algorithm. The `reserve_policy` attribute determines the type of reserve methodology the device driver will implement when the device is opened, and it can be used to limit device access from other adapters, whether on the same system or another system. The algorithm attribute defines the methodology that the PCM uses to manage I/O across the paths configured for a device.

Supported multi-path devices

The AIX default PCMs support a set of disk devices, including Symmetrix devices, defined in the `devices.common.IBM.mpio.rte` fileset.

IMPORTANT

A Symmetrix device configured as an **MPIO other FC device** is **not** supported.

The only supported configuration is with the EMC Symmetrix FCP MPIO ODM predefined attributes which has been certified to operate with the AIX default PCMs. These predefines get installed with the `EMC.Symmetrix.fcp.MPIO.rte` fileset.

MPIO device attributes

This section describes attributes that are available only by multi-path devices. The attributes can be displayed or changed using the commands (in particular, the **`lsattr`** and the **`chdev`** commands).

The required device attributes that all MPIO devices must support is `reserve_policy`. Typically, a multi-path device also has the `PR_key` device attribute. A multi-path device can have additional device- specific attributes.

Other device-specific attributes are as follows:

- `reserve_policy`

  Defines whether a reservation methodology is employed when the device is opened. The values are as follows:

  - **`no_reserve`**
    Does not apply a reservation methodology for the device. The device might be accessed by other initiators, and these initiators might be on other host systems.
  - **`single_path`**
    Applies a SCSI2 reserve methodology for the device, which means the device can be accessed only by the initiator that issued the reserve. This policy, which is the default setting for Symmetrix devices, prevents other initiators in the same host or on other hosts from accessing the device. This policy uses the SCSI2 reserve to lock the device to a single initiator (path), and commands routed through any other path result in a reservation conflict.
  - **`PR_exclusive`**
    Applies a SCSI3 persistent-reserve, exclusive-host methodology when the device is opened. The `PR_key` attribute value must be unique for each host system. The `PR_key` attribute is used to prevent access to the device by initiators from other host systems. This setting is not supported by EMC at this time.
  - **`PR_shared`**
    Applies a SCSI3 persistent-reserve, shared-host methodology when the device is opened. The `PR_key` value must be a unique value for each host system. Initiators from other host systems are required to register before they can access the device. This setting is only supported with PowerMax 8000, PowerMax 2000, VMAX All Flash 950F/950FX, VMAX All Flash 250F/250FX/450F/450FX/850F/850FX,
  - **`PR_key`**

  Required only if the device supports any of the persistent reserve policies (`PR_exclusive` or `PR_shared`). `PR_shared` is only supported with PowerMax 8000, PowerMax 2000, VMAX All Flash 950F/950FX, VMAX All Flash 250F/250FX/450F/450FX/850F/850FX.

Path control module attributes

The following are the device attributes for the AIX default PCMs:

- `algorithm`

  Determines the methodology by which the I/O is distributed across the paths for a device. The algorithm attribute has the following value:

  - **`failover`**

    Sends all I/O down a single path. If the path is determined to be faulty, an alternate path is selected for sending all I/O. This algorithm keeps track of all the enabled paths in an ordered list. If the path being used to send I/O becomes marked failed or disabled, the next enabled path in the list is selected. The sequence within the list is determined by the path priority “path priority” attribute.

- round_robin

  Distributes the I/O across all enabled paths (load balancing). The path priority is determined by the path priority "path priority" attribute value. If a path becomes marked failed or disabled, it is no longer used for sending I/O. The priority of the remaining paths

is then recalculated to determine the percentage of I/O that should be sent down each path. If all paths have the same value, then the I/O is distributed equally across all enabled paths. This algorithm is available only if the **reserve_policy** is set to **no_reserve**.

- **hcheck_mode**

  Determines which paths should be checked when the health check capability is used. The attribute supports the following modes:

  - **enabled**
    Sends the **healthcheck** command down paths with a state of enabled.
  - **failed**

    Sends the **healthcheck** command down paths with a state of failed.
  - **nonactive**

    (Default) Sends the **healthcheck** command down paths that have no active I/O, including paths with a state of failed. If the algorithm selected is failover, then the **healthcheck** command is also sent on each of the paths that have a state of enabled but have no active I/O. If the algorithm selected is **round_robin**, then the **healthcheck** command is only sent on paths with a state of failed, because the **round_robin** algorithm keeps all enabled paths active with I/O.

- **hcheck_interval**

  Defines how often the health check is performed on the paths for a device. The attribute supports a range from 0 to 3600 seconds. When a value of 0 is selected, health checking is disabled.

- **dist_tw_width**

  Defines the duration of a time window. This is the time frame during which the distributed error detection algorithm will cumulate I/Os returning with an error. The **dist_tw_width** attribute unit of measure is milliseconds. Lowering this attributes value decreases the time duration of each sample taken and decreases the algorithms sensitivity to small bursts of I/O errors. Increasing this attribute's value will increase the algorithms sensitivity to small bursts of errors and the probability of failing a path.

- **dist_err_percent**

  Defines the percentage of time windows having an error allowed on a path before the path is failed due to poor performance. The **dist_err_percent** has a range from 0-100. The distributed error detection algorithm will be disabled when the attribute is set to zero (0). The default setting is zero. The distributed error detection algorithm samples the fabric connecting the device to the adapter for errors. The algorithm calculates a percentage of samples with errors and will fail a path if the calculated value is larger than the **dist_err_percent** attribute value.

- **path priority**

  Modifies the behavior of the algorithm methodology on the list of paths. When the algorithm attribute value is failover, the paths are kept in a list. The sequence in this list determines which path is selected first and, if a path fails, which path is selected next. The sequence is determined by the value of the path priority attribute. A priority of 1 is the highest priority. Multiple paths can have the same priority value, but if all paths have the same value, selection is based on when each path was configured.

When the algorithm attribute value is round_robin, the sequence is determined by percent of I/O. The path priority value determines the percentage of the I/O that should be processed down each path. I/O is distributed across the enabled paths. A path is selected until it meets its required percentage. The algorithm then marks that path failed or disabled to keep the distribution of I/O requests based on the path priority value.

Supported multiple path I/O configurations

Dell EMC supports MPIO connectivity in a heterogeneous configurations with EMC Symmetrix, VNX series, or CLARiiON storage systems and IBM storage arrays only. At this time, the following are not supported:

- IBM SDD and IBM RDAC connectivity to Dell EMC storage arrays

The following configurations are supported:

- Minimum AIX 5.2 ML-02 with minimum APAR IY49586.
- Minimum EMC ODM version 5.2.0.0 required which contains the Minimum EMC ODM version 5.2.0.0 required which contains the EMC.Symmetrix.fcp.MPIO.rte' fileset. The ODM package can be downloaded from the ftp server `ftp://ftp.emc.com/pub/elab/aix/ODM_DEFINITIONS` directory. Read the `README.1st` file, in the ODM_DEFINITIONS directory, carefully before upgrading'EMC.Symmetrix.fcp.MPIO.rte' fileset. The ODM package can be downloaded from the EMC ftp server `ftp://ftp.emc.com/pub/elab/aix/ODM_DEFINITIONS` directory. Please read the `README.1st` file, in the ODM_DEFINITIONS directory, carefully before upgrading.
- Minimum AIX 5.2 ML-02 with minimum APAR IY49586.
- Minimum EMC ODM version 5.2.0.0 required which contains the 'EMC.Symmetrix.fcp.MPIO.rte' fileset. The ODM package can be downloaded from the EMC ftp server `ftp://ftp.emc.com/pub/elab/aix/ODM_DEFINITIONS` directory. Please read the `README.1st` file, in the ODM_DEFINITIONS directory, carefully before upgrading.
- IBM native Fibre Channel driver support only.
- MPIO is not supported with the Symmetrix 3000 or Symmetrix 5000.
- MPIO support for Symmetrix includes SCSI-2 reserve support with the hdisk device attribute set to `reserve_policy=single_path`. Device reserve_policy `no_reserve` is supported by Dell EMC when load balancing is required. Device values `PR_shared` is supported with PowerMax 8000, PowerMax 2000, VMAX All Flash 950F/950FX, VMAX All Flash 250F/250FX/450F/450FX/850F/850FX.
- Coexistence configuration support of EMC Symmetrix storage arrays and IBM storage arrays into MPIO on the same AIX image is supported with the minimum following AIX APARs:

  IY49013 – AIX Maintenance Level 03.

  IY56376 – Fix for reboot loop on machines with many devices.

  IY56769 – Latest AIX 5.2 updates as of May 2004.

The following combinations of MPIO are supported:

- **Symmetrix-MPIO + VNX series or CLARiiON-PowerPath** — All Symmetrix devices must be exclusively configured into MPIO and all CLARiiON devices must be exclusively configured into PowerPath with minimum version 3.0.4.2. No Symmetrix devices can be configured into PowerPath. (i.e. devices under MPIO control can not also be configured into PowerPath at this time).
- **Symmetrix-PowerPath + VNX series or CLARiiON MPIO**— Dedicated HBAs for Symmetrix devices configured into PowerPath and dedicated HBAs for CLARiiON devices configured into MPIO.
- **Symmetrix-PowerPath + IBM Storage-MPIO** — Dedicated HBAs for Symmetrix devices configured into PowerPath and dedicated HBAs for IBM storage devices into MPIO.
- **Symmetrix-MPIO + IBM Storage-MPIO** — Shared HBAs for Symmetrix devices configured into MPIO and shareed HBAs for IBM storage devices into MPIO.
- **Symmetrix-MPIO + IBM FAStT-RDAC** — Dedicated HBAs for Symmetrix devices configured into MPIO and dedicated HBAs for IBM FAStT storage devices into RDAC.

- **Symmetrix-MPIO + VNX series or CLARiiON-PowerPath + IBM Storage-MPIO**— Shared HBAs for Symmetrix devices configured into MPIO and dedicated HBAs for the CLARiiON devices configured into PowerPath and shared HBAs for IBM storage arrays configured into MPIO.
- **Symmetrix-MPIO + FAStT-RDAC + IBM Storage-MPIO** — Shared HBAs for Symmetrix devices configured into MPIO and dedicated HBAs for the IBM FAStT storage devices configured into RDAC and shared HBAs for IBM storage arrays configured into MPIO.

Fast I/O failure for Fibre Channel devices

As of AIX release 5.2 Maintenance level 2 (APAR IY48488), IBM has introduced a new feature called *Fast I/O Failure*. If the Fibre Channel adapter driver detects a link event, such as a lost link between a storage device and a switch, the driver waits approximately 15 seconds to allow the fabric to stabilize. After this wait period, the driver detects that the device is not on the fabric and begins failing all I/Os. Any new I/Os are failed immediately by the adapter until the driver detects that the device has returned.

Fast I/O Failure is controlled by the new attribute `fc_err_recov` in the `fscsi` that is defined in the ODM. The default setting of this attribute is `delayed_fail`, which in effect disables this feature. Setting the attribute to `fast_fail` enables the Fast Failover feature.

Here is an example of enabling the feature on the `fscsi0` adapter:

```
chdev -l fscsi0 -a fc_err_recov=fast_fail -P
```

The above command would have to be repeated for all present `fscsi` adapters in the system.

The Fast Fail feature is now enabled for the `fscsi0` adapter. When the adapter driver receives an event from the switch that there has been a link event (RSCN), the driver invokes the Fast Fail logic.

Fast Fail functionality is desirable when multipathing software (such as PowerPath) is installed. Setting the `fc_err_recov` attribute to `fast_fail` should decrease the failure times due to a link failures.

In a single direct-connect path environment, it is recommended that the Fast Fail feature be disabled.

Requirements

The requirements for Fast I/O Failure support are:

- Fast I/O Failure is supported only in a switched environment. It is not supported in arbitrated loop environments.
- FC 6227 adapter firmware — level 3.22A1 or greater.
- FC 6228 adapter firmware — level 3.82a1 or greater.
- FC 6239 adapter firmware — level 1.00X5 or greater.
- FC 5716 adapter firmware — level 1.90A4 or greater.
- FC 1977 adapter firmware — level 1.90A4 or greater.
- Newer released HBAs contain base support.

Failure to meet these requirements can cause the `fscsi` device to log an error log of type `INFO` indicating that one of these requirements was not met and that fast fail *has not* been enabled.

Dynamic tracking of Fibre Channel devices

As of AIX release 5.2 Maintenance level 2 (APAR IY48488), IBM has introduced a new feature called *Dynamic Tracking of Fibre Channel Devices*. Previous versions of the operating system required an administrator to delete hdisks and adapters before making changes to the fabric that would result in a different `N_Port ID`. For example, if moving a fibre cable from one switch port to another, a device reconfiguration would be required.

If Dynamic Tracking is enabled, the Fibre Channel adapter driver will detect the occurrence of a change in the `N_Port ID` and automatically reroute I/Os for that device to the new address. Examples of events that can cause an `N_Port ID` to change are moving a cable between a switch and storage device from one switch port to another, connecting two separate switches via an interswitch link (ISL), and possibly rebooting a switch.

The Dynamic Tracking for Fibre Channel devices is controlled by a new fscsi attribute called `dyntrk`. The default setting for this attribute is **no** (disabled), Setting this attribute to **yes** enables the feature for the named adapter.

The following example enables Dynamic Tracking for adapter `fscsi0`:

```
chdev -l fscsi0 -a dyntrk=yes -P
```

Requirements

Requirements for Dynamic Tracking support are:

- Fast Fail is supported only in a switched environment. It is not supported in arbitrated loop environments.
- FC 6227 adapter firmware — level 3.22A1 or greater.
- FC 6228 adapter firmware — level 3.82a1 or greater.
- FC 6239 adapter firmware — level 1.00X5 or greater.
- FC 5716 adapter firmware — level 1.90A4 or greater.
- FC 1977 adapter firmware — level 1.90A4 or greater.
- Newer released HBAs contain base support.
- The device World Wide Name (Port Name) and Node Name must remain constant, and the World Wide Name must be unique. Changing the World Wide Name or Node Name of an available or online device could result in I/O failures.
- Each Fibre Channel storage device instance must have `world_wide_name` and `node_name` attributes, both of which are included in EMC AIX ODM package 5.1.0.0 and later.

  Updated filesets that contain the `sn_location` attribute (discussed in the list item) should also be updated to contain the `world_wide_name` and `node_name`.

- The storage device must provide a reliable method to extract a unique serial number for each LUN. The AIX Fibre Channel device drivers will not autodetect serial number location, so the method for serial number extraction must be explicitly provided by any storage vendor in order to support dynamic tracking for their devices.

  This information is conveyed to the drivers using the `sn_location` ODM attribute for each storage device. If the disk or tape driver detects that the **`sn_location`** ODM attribute is missing, an error log of type `INFO` will be generated and dynamic tracking *will not* be enabled.

  Note: The `sn_location` attribute may be non-displayable, so running the **`lsattr`** command on an `hdisk`, for example, may not show the attribute, but it may, indeed, be present in ODM.

- The Fibre Channel device drivers will be able to track devices on a SAN fabric (a fabric as seen from a single host bus adapter) if the `N_Port IDs` on the fabric stabilize within about 15 seconds. If cables are not reseated or `N_Port IDs` continue to change after the initial 15 seconds, I/O failures could result.
- Devices will not be tracked across host bus adapters. Devices will only track if they remain visible from the same HBA that they were originally connected to.

  For example, if device A were moved from one location to another on fabric A attached to host bus adapter A (that is, its `N_Port on fabric A` changes), the device would seamlessly be tracked without any user intervention and I/O to this device can continue.

  However, if a device A is visible from HBA A, but not from HBA B, and device A is moved from the fabric attached to HBA A to the fabric attached to HBA B, device A will not be accessible on fabric A nor on fabric B. User intervention would be required to make it available on fabric B by invoking `cfgmgr`. The AIX device instance on fabric A would no longer be usable, and a new device instance on fabric B would be created. This device would have to manually be added to volume groups, multipath device instances, etc. In essence, this is the same as removing a device from fabric A and adding a new device to fabric B.
- No dynamic tracking will performed for Fibre Channel dump devices while an AIX system dump is in progress. In addition, dynamic tracking is not supported during boot or during cfgmgr invocations. SAN changes should not be made while any of these operations are in progress.
- Once devices are tracked, ODM will potentially contain stale information as SCSI IDs in ODM will no longer reflect actual SCSI IDs on the SAN. ODM will remain in this state until cfgmgr is run manually or the system is rebooted, provided all drivers, including any third-party Fibre Channel SCSI target drivers, are dynamic-tracking capable. If **`cfgmgr`** is run manually, **`cfgmgr`** must be invoked on all affected `fscsi` devices, which can easily be accomplished by running `cfgmgr` without any options, or by invoking **`cfgmgr`** on each `fscsi` device individually.

  Note: Running `cfgmgr` at run time to recalibrate the SCSI IDs may not update the SCSI ID in ODM for a storage device if the storage device is currently opened, such as when volume groups are varied on. **`cfgmgr`** would need to be run on devices that are not opened or the system should be rebooted to recalibrate the SCSI IDs. Note that stale SCSI IDs in ODM have no adverse affect on the FC drivers and recalibration of SCSI IDs in ODM is not necessary for the FC drivers to function properly. Any applications that

communicate with the adapter driver directly via ioctl calls and use the SCSI ID values from ODM, however, need to be updated as indicated in the next bullet to avoid using potentially stale SCSI IDs.

- Even with dynamic tracking enabled users are strongly encouraged to make SAN changes, such as cable moves/swaps and establishing ISL links, during maintenance windows. Making SAN changes during full production runs is discouraged. This is due to the fact that there is a short interval of time to perform any SAN changes. Cables that are not reseated correctly, for example, could result in I/O failures. Performing these operations during a time of little/no traffic minimizes impact of I/O failures due to misplugging of cables, taking too long to re-cable, and so forth.

Fast Fail and Dynamic Tracking interaction

Although Fast Fail and Dynamic Tracking of Fibre Channel devices are technically separate features, the enabling of one may change the interpretation of the other in certain situations. The following table shows the behavior exhibited by the Fibre Channel drivers with the various permutations of these settings.

dyntrk	fc_err_recov	Fibre Channel Driver Behavior
no	`delayed_fail`	This is the default setting. This is legacy behavior existing in previous versions of AIX. The FC drivers will not recover if the `scsi_id` of a device changes, and I/Os will take longer to fail when a link loss occurs between a remote storage port and switch. This may be desirable in single-path situations if dynamic tracking support is not a requirement.
no	`fast_fail`	If the driver receives a Registered State Change Notification (RSCN) from the switch, this could indicate a link loss between a remote storage port and switch. After an initial 15 second delay, the FC drivers will query to see if the device is on the fabric. If not, I/Os will be flushed back by the adapter. Future retries or new I/Os will fail immediately if the device is still not on the fabric. If the FC drivers detects the device is on the fabric but the scsi_id has changed, the FC device drivers will not recover, i.e., I/Os will be failed with PERM errors.
yes	`delayed_fail`	If the driver receives a Registered State Change Notification (RSCN) from the switch, this could indicate a link loss between a remote storage port and switch. After an initial 15 second delay, the FC drivers will query to see if the device is on the fabric. If not, I/Os will be flushed back by the adapter. Future retries or new I/Os will fail immediately if the device is still not on the fabric, although the storage driver (disk, tape) drivers may inject a small delay (2-5 seconds) between I/O retries. If the FC drivers detects the device is on the fabric but the scsi_id has changed, the FC device drivers will reroute traffic to the new `scsi_id`.

dyntrk	fc_err_recov	Fibre Channel Driver Behavior (continued)
yes	`fast_fail`	If the driver receives a Registered State Change Notification (RSCN) from the switch, this could indicate a link loss between a remote storage port and switch. After an initial 15 second delay, the FC drivers will query to see if the device is on the fabric. If not, I/Os will be flushed back by the adapter. Future retries or new I/Os will fail immediately if the device is still not on the fabric. The storage driver (disk, tape) will likely not delay between retries. If the FC drivers detects the device is on the fabric but the `scsi_id` has changed, the FC device drivers will reroute traffic to the new `scsi_id`.

Note that with dynamic tracking disabled, there is a marked difference between the `delayed_fail` and `fast_fail` settings of the `fc_err_recov` attribute. However, with dynamic tracking enabled, the setting of the `fc_err_recov` attribute is less significant. This is because there is some overlap in the dynamic tracking and fast fail error recovery policies. As such, enabling dynamic tracking inherently enables some of the fast fail logic.

The general error recovery procedure when a device is no longer reachable on the fabric is the same for both `fc_err_recov` settings with dynamic tracking enabled, with the minor difference that the storage drivers may choose to inject delays between I/O retries if `fc_err_recov` is set to `delayed_fail`. This will increase the I/O failure time by some additional amount depending on the delay value and number of retries before permanently failing the I/O. With high I/O traffic, however, the difference between `delayed_fail` and `fast_fail` may be more noticeable.

SAN administrators may want to experiment with these settings to find the correct combination of settings for their environment.

VMAX and VMAX3 outputs with the lscfg command

This section describes the different outputs of the **lscfg** command for VMAX and VMAX3 storage arrays. The information will help EMC Customer Support to understand the reason for the different outputs.

To use this information, you need a working knowledge of AIX operating system commands and EMC Symmetrix and/or CLARiiON storage arrays.

This information applies to systems with the IBM native Fibre Channel driver using feature code 6227, 6228, or 6239 HBAs that are connected to Symmetrix or to a Symmetrix array. It includes the VMAX3 family, VMAX, and DMX series storage arrays running Dell EMC FLARE™ release 11 (02.04 for CX series or 8.49 for FC4700 series) or release 12 (02.05 for CX series or 8.50 for FC4700 series) and/or Symmetrix storage arrays.

Background information

When you install the ODM version 5.3.1.0 definitions on AIX and run the **`lscfg -vl hdiskNN`** command to inquire about disk information, the outputs are different for VMAX3 and VMAX arrays. The output shows only a portion of the disk information with VMAX3, but there is more disk information in the output when connected to a VMAX array.

VMAX3 output

```
#lscfg -vl hdisk628
hdisk628
U2C4E.001.DBJ9390-P2-C6-T2-W50000973580DA844-L1000000000
000 EMC Symmetrix FCP MPIO VG3R
Manufacturer.............EMC
Machine Type and Model......SYMMETRIX
ROS Level and ID............5977
Device
Specific.(Z0).........600009700BB96C09146D003200000006
```

VMAX output

```
#lscfg -vl hdisk38
hdisk38
U8231.E2D.0668ADT-V6-C4-T1-W5000097300249DA4-LA000000000
000 EMC Symmetrix FCP MPIO VRAID
Manufacturer.............EMC
Machine Type and Model...... SYMMETRIX
ROS Level and ID............5876
Serial Number.............. 435E6000
Part Number ......00000000000400969000195
EC Level......702343
LIC Node VPD.............05E6
Device Specific.(Z0).........00
Device Specific.(Z1).........40
Device Specific.(Z2)........ 000000300D9C09000260019C
Device Specific.(Z3)........ 12000000
```

```
Device Specific.(Z4)........ 54110000
Device Specific.(Z5)........ 4F80
Device Specific. (Z6)......... 45
```

The reason for this difference

After further investigation, we determined that the discrepancy in output for the **`lscfg`** command between VMAX and VMAX3 is the result of changes imposed on the ODM kit by VMAX3 architecture. Here is some of the background:

- The data returned for compatibility volumes in VMAX2 was obtained from the inquiry **page 0** command.
- When VMAX3 was designed, support for volume counts above 64k and mobility devices was included.
- For these new device features, a significant portion of the data originally returned by the **`lscfg`** command (meta volumes, RAID5, etc.) from the **page 0** command is no longer relevant.
- VMAX Engineering placed device data in new locations other than **page 0** and requires all inquiry requests for all device types, including both legacy compatibility and mobility devices, to be made from these new locations; the information presented back to the host must be the same for both compatibility and mobility volumes.

Because of these design restrictions, the best explanation is that VMAX3 architecture imposed some of the changes customers are seeing. If customers can identify the specific data that they need to manage their AIX environments, we can probably find another method for obtaining it; other AIX commands or **`SymmCLI`** are possible methods.

CHAPTER 7

AIX and iSCSI

This chapter provides a brief overview of AIX iSCSI attached to EMC storage.

Note: Any general references to Symmetrix or Symmetrix array include the VMAX3 Family, VMAX, and DMX.

Getting started

This section includes the following information to help you get started:

Host configuration

In an AIX iSCSI environment, you are required to install the specific iSCSI ODM filesets when connecting to EMC VMAX400K, VMAX 200K, VMAX 100K, VMAX 40K, VMAX 20K/VMAX, VMAX 10K (Systems with SN xxx987xxxx), VMAX 10K (Systems with SN xxx959xxxx), VMAXe, VNX series, CLARiiON, or Celerra storage arrays.

Note: VNX iSCSI ODM filesets are required for VNXe 3200 storage arrays

Note: Celerra ODM filesets are required for VNXe 3150 and VNXe 3300 storage arrays.

- AIX 7.1 requires minimum ODM kit 6.0.0.5.
- AIX 6.1 requires minimum ODM kit 5.3.1.0

Installing the ODM support package

Perform the following procedure if you are installing any of the following iSCSI filesets:

- Symmetrix iSCSI Support Software
- CLARiiON iSCSI Support Software
- Celerra iSCSI Support Software

1. Log on to the AIX server as root.
2. Obtain the ODM package from the FTP server:
   - **ftp ftp.emc.com x**
   - **cd /pub/elab/aix/ODM_DEFINITIONS x**
   - **bin x**
   - **get EMC.AIX.5.3.0.2.tar.Z (AIX 5.2, 5.3 or 6.1)**
   - **bye**

   Note: These are the latest filesets as of this printing. If a more recent fileset is available, use it.

3. In the `/tmp` directory, uncompress the tar package:

   ```
   uncompress EMC.AIX.5.3.0.1.tar.Z
   ```
4. Extract the tar package:

```
tar -xvf EMC.AIX.5.3.0.1.tar
```

Using SMIT

To complete the installation using SMIT:

1. At the command line, invoke `SMIT` from the `/tmp` directory:

   `smitty installp`

2. Select **Install and Update from Latest Available Software**.
3. Select the input device to be the current working directory:

   ./

4. Select **F4=List** to list available software.
5. Select **F7=Select** the appropriate iSCSI fileset.
6. Select **Enter** to perform the action.
7. Accept the default options and press **Enter** to initiate the installation.
8. Scroll to the bottom of the installation summary screen to verify that the SUCCESS message is displayed.
9. Exit SMIT.

The AIX iSCSI software initiator uses the `/etc/iscsi/targets` file to identify and configure iSCSI targets. The AIX config-method script (cfgmgr) will search the `targets` file for a valid iSCSI target. The `targets` file should contain an IP address, followed by the TCP port identifier (3260) and the iSCSI Qualified Name (`iqn`) of the target.

Example Targets file:

```
# Symmetrix
11.2.6.20 3260 iqn.1992-04.com.emc:5006048ad5f00d01
# Clariion
11.2.6.21 3260 iqn.1992-04.com.emc:cx.hk192201153.a1
```

The **`lsattr -El iscsi0`** command will display the iqn name for the iscsi0 driver.

The AIX software initiator uses the iscsi0 device driver. IBM NIC cards are configured through the `enX` interface.

Example **`ifconfig -a`** output:

```
# ifconfig -a
en2: flags=1e080863,c0 <UP,BROADCAST,NOTRAILERS,RUNNING,SIMPLEX,MULTICAST,GROUPRT
,64BIT,CHECKSUM_OFFLOAD(ACTIVE),CHAIN> inet 11.2.5.141 netmask 0xffffff00 broadcast 11.2.5.255
tcp_sendspace 131072 tcp_recvspace 65536
en3: flags=1e080863,c0 <UP,BROADCAST,NOTRAILERS,RUNNING,SIMPLEX,MULTICAST,GROUPRT
,64BIT,CHECKSUM_OFFLOAD(ACTIVE),CHAIN> inet 11.2.6.141 netmask
0xffffff00 broadcast 11.2.6.255 tcp_sendspace 131072 tcp_recvspace 65536
```

Symmetrix and iSCSI

The Symmetrix requires LUN masking for iSCSI attach.

- For DMX-4 and earlier arrays, enable the VCM director bit on all SE directors attached to AIX iSCSI initiators.
- For VMAX 20K/VMAX, enable the ACLX director bit on all SE directors attached to AIX iSCSI initiators.

The full iqn name of the iSCSI initiator will be used when assigning LUNs via LUN masking.

The Symmetrix supports jumbo frames up to 9000 mtu. Check with IBM to verify if the Network Interface Card supports jumbo frames.

Always refer to the Dell EMC Simple Support Matrices for the most up-to-date support information.

Multipath I/O

EMC currently supports PowerPath 5.1 or later for iSCSI multipath functionality with AIX.

Note: Dell EMC does *not* currently support AIX MPIO.

Always refer to the Dell EMC Simple Support Matrices for the most up-to-date support information.

Clusters

HACMP and GPFS are currently not supported on AIX servers iSCSI attached to Symmetrix arrays.

CHAPTER 8

Virtual Provisioning

This chapter provides information about Virtual Provisioning and IBM AIX.

Note: For further information regarding the correct implementation of Virtual Provisioning, refer to the *Symmetrix Virtual Provisioning Implementation and Best Practices Technical Note*, available at Dell EMC Online Support.

Note: Any general references to Symmetrix or Symmetrix array include the VMAX3 Family, VMAX, and DMX.

Virtual Provisioning on Symmetrix

Dell EMC Virtual Provisioning™ enables organizations to improve speed and ease of use, enhance performance, and increase capacity utilization for certain applications and workloads. Dell EMC Symmetrix Virtual Provisioning integrates with existing device management, replication, and management tools, enabling customers to easily build Virtual Provisioning into their existing storage management processes.

Virtual Provisioning, which marks a significant advancement over technologies commonly known in the industry as “thin provisioning,” adds a new dimension to tiered storage in the array, without disrupting organizational processes.

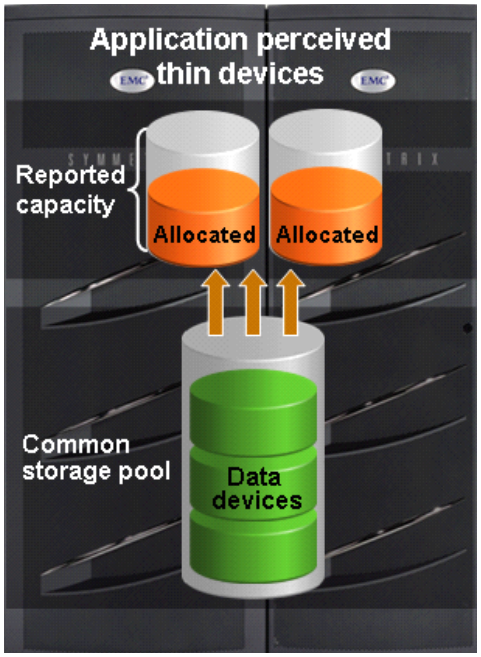


Figure 10 Virtual provisioning on Symmetrix

Terminology

This section provides common terminology and definitions for Symmetrix and thin provisioning.

Symmetrix

Basic Symmetrix terms include:

Device	A logical unit of storage defined within an array
Device capacity	The storage capacity of a device
Device extent	Specifies a quantum of logically contiguous blocks of storage
Host accessible device	A device that can be made available for host use
Internal device	A device used for a Symmetrix internal function that cannot be made accessible to a host
Storage pool	A collection of internal devices for some specific purpose

Thin provisioning

Basic thin provisioning terms include:

Thin device	A host accessible device that has no storage directly associated with it
Data device	An internal device that provides storage capacity to be used by thin devices
Thin pool	A collection of data devices that provide storage capacity for thin devices
Thin pool capacity	The sum of the capacities of the member data devices
Thin pool allocated capacity	A subset of thin pool enabled capacity that has been allocated for the exclusive use of all thin devices bound to that thin pool
Thin device user pre-allocated capacity	The initial amount of capacity that is allocated when a thin device is bound to a thin pool. This property is under user control
Bind	Refers to the act of associating one or more thin devices with a thin pool
Pre-provisioning	An approach sometimes used to reduce the operational impact of provisioning storage. The approach consists of satisfying provisioning operations with larger devices that needed initially, so that the future cycles of the storage provisioning process can be deferred or avoided

Over-subscribed thin pool	A thin pool whose thin pool capacity is less than the sum of the reported sizes of the thin devices using the pool
Thin device extent	The minimum quantum of storage that must be mapped at a time to a thin device
Data device extent	The minimum quantum of storage that is allocated at a time when dedicating storage from a thin pool for use with a specific thin device

Thin device

Symmetrix Virtual Provisioning introduces a new type of host-accessible device called a *thin device* that can be used in many of the same ways that regular host-accessible Symmetrix devices have traditionally been used. Unlike regular Symmetrix devices, thin devices do not need to have physical storage completely allocated at the time the devices are created and presented to a host. The physical storage that is used to supply disk space for a thin device comes from a shared thin storage pool that has been associated with the thin device.

A thin storage pool is comprised of a new type of internal Symmetrix device called a *data device* that is dedicated to the purpose of providing the actual physical storage used by thin devices. When they are first created, thin devices are not associated with any particular thin pool. An operation referred to as *binding* must be performed to associate a thin device with a thin pool.

When a write is performed to a portion of the thin device, the Symmetrix allocates a minimum allotment of physical storage from the pool and maps that storage to a region of the thin device, including the area targeted by the write. The storage allocation operations are performed in small units of storage called *data device extents*. A round-robin mechanism is used to balance the allocation of data device extents across all of the data devices in the pool that have remaining un-used capacity.

When a read is performed on a thin device, the data being read is retrieved from the appropriate data device in the storage pool to which the thin device is bound. Reads directed to an area of a thin device that has not been mapped does not trigger allocation operations. The result of reading an unmapped block is that a block in which each byte is equal to zero will be returned. When more storage is required to service existing or future thin devices, data devices can be added to existing thin storage pools. New thin devices can also be created and associated with existing thin pools.

It is possible for a thin device to be presented for host use before all of the reported capacity of the device has been mapped. It is also possible for the sum of the reported capacities of the thin devices using a given pool to exceed the available storage capacity of the pool. Such a thin device configuration is said to be *over-subscribed*.

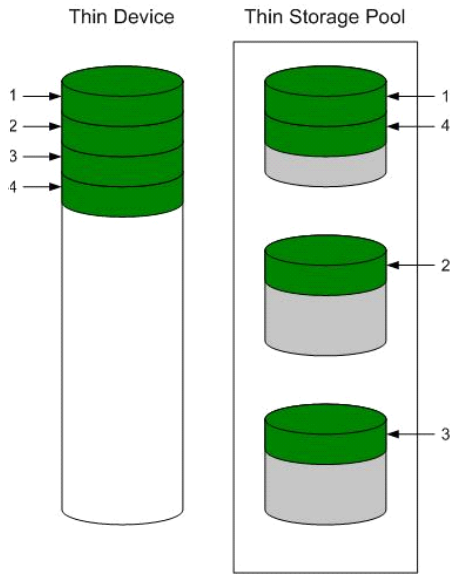


Figure 11 Thin device and thin storage pool containing data devices

In Figure 11, as host writes to a thin device are serviced by the Symmetrix array, storage is allocated to the thin device from the data devices in the associated storage pool. The storage is allocated from the pool using a round-robin approach that tends to stripe the data devices in the pool.

Implementation considerations

When implementing Virtual Provisioning, it is important that realistic utilization objectives are set. Generally, organizations should target no higher than 60 percent to 80 percent capacity utilization per pool. A buffer should be provided for unexpected growth or a "runaway" application that consumes more physical capacity than was originally planned for. There should be sufficient free space in the storage pool equal to the capacity of the largest unallocated thin device.

Organizations also should balance growth against storage acquisition and installation timeframes. It is recommended that the storage pool be expanded before the last 20 percent of the storage pool is utilized to allow for adequate striping across the existing data devices and the newly added data devices in the storage pool.

Thin devices can be deleted once they are unbound from the thin storage pool. When thin devices are unbound, the space consumed by those thin devices on the associated data devices is reclaimed.

Note: Users should first replicate the data elsewhere to ensure it remains available for use.

Data devices can also be disabled and/or removed from a storage pool. Prior to disabling a data device, all allocated tracks must be removed (by unbinding the associated thin devices). This means that all thin devices in a pool must be unbound before any data devices can be disabled.

This section contains the following information:

Over-subscribed thin pools

It is permissible for the amount of storage mapped to a thin device to be less than the reported size of the device. It is also permissible for the sum of the reported sizes of the thin devices using a given thin pool to exceed the total capacity of the data devices comprising the thin pool. In this case the thin pool is said to be *over-subscribed*. Over-subscribing allows the organization to present larger-than-needed devices to hosts and applications without having to purchase enough physical disks to fully allocate all of the space represented by the thin devices.

The capacity utilization of over-subscribed pools must be monitored to determine when space must be added to the thin pool to avoid out-of-space conditions.

Not all operating systems, filesystems, logical volume managers, multipathing software, and application environments will be appropriate for use with over-subscribed thin pools. If the application, or any part of the software stack underlying the application, has a tendency to produce dense patterns of writes to all available storage, thin devices will tend to become fully allocated quickly. If thin devices belonging to an over-subscribed pool are used in this type of

environment, out-of-space and undesired conditions may be encountered before an administrator can take steps to add storage capacity to the thin data pool. Such environments are called *thin-hostile.*

Thin-hostile environments

There are a variety of factors that can contribute to making a given application environment thin-hostile, including:

- One step, or a combination of steps, involved in simply preparing storage for use by the application may force all of the storage that is being presented to become fully allocated.
- If the storage space management policies of the application and underlying software components do not tend to reuse storage that was previously used and released, the speed in which underlying thin devices become fully allocated will increase.
- Whether any data copy operations (including disk balancing operations and de-fragmentation operations) are carried out as part of the administration of the environment.
- If there are administrative operations, such as bad block detection operations or file system check commands, that perform dense patterns of writes on all reported storage.
- If an over-subscribed thin device configuration is used with a thin-hostile application environment, the likely result is that the capacity of the thin pool will become exhausted before the storage administrator can add capacity unless measures are taken at the host level to restrict the amount of capacity that is actually placed in control of the application.

Pre-provisioning with thin devices in a thin hostile environment

In some cases, many of the benefits of pre-provisioning with thin devices can be exploited in a thin-hostile environment. This requires that the host administrator cooperate with the storage administrator by enforcing restrictions on how much storage is placed under the control of the thin-hostile application.

For example:

- The storage administrator pre-provisions larger than initially needed thin devices to the hosts, but only configures the thin pools with the storage needed initially. The various steps required to create, map, and mask the devices and make the target host operating systems recognize the devices are performed.
- The host administrator uses a host logical volume manager to carve out portions of the devices into logical volumes to be used by the thin-hostile applications.
- The host administrator may want to fully preallocate the thin devices underlying these logical volumes before handing them off to the thin-hostile application so that any storage capacity shortfall will be discovered as quickly as possible, and discovery is not made by way of a failed host write.
- When more storage needs to be made available to the application, the host administrator extends the logical volumes out of the thin devices that have already been presented. Many databases can absorb an additional disk partition non-disruptively, as can most file systems and logical volume managers.
- Again, the host administrator may want to fully allocate the thin devices underlying these volumes before assigning them to the thin-hostile application.

  In this example it is still necessary for the storage administrator to closely monitor the over-subscribed pools. This procedure will not work if the host administrators do not observe restrictions on how much of the storage presented is actually assigned to the application.

Host boot/root/swap/dump devices positioned on Symmetrix VP (tdev) devices

A boot /root /swap /dump device positioned on Symmetrix Virtual Provisioning (thin) device(s) is supported with Enginuity 5773 and later. However, some specific processes involving boot /root/swap/dump devices positioned on thin devices should not have exposure to encountering the out-of-space condition. Host-based processes such as kernel rebuilds, swap, dump, save crash, and Volume Manager configuration operations can all be affected by the thin provisioning out-of-space condition. This exposure is not specific to Dell EMC's implementation of thin provisioning. Dell EMC strongly recommends that the customer avoid encountering the out-of-space condition involving boot / root /swap/dump devices positioned on Symmetrix VP (thin) devices using the following recommendations;

- It is strongly recommended that Virtual Provisioning devices utilized for boot /root/dump/swap volumes must be fully allocated[1] or the VP devices must not be oversubscribed[2].

  Should the customer use an over-subscribed thin pool, they should understand that they need to take the necessary precautions to ensure that they do not encounter the out-of-space condition.

1. A fully allocated Symmetrix VP (thin) device has 100% of the advertised space mapped to blocks in the data pool that it is bound to. This can be achieved by use of the Symmetrix VP pre-allocation mechanism or host-based utilities that will enforce pre-allocation of the space (such as, host device format.)
2. An over-subscribed Symmetrix VP (thin) device is a thin device, bound to a data pool, that does not have sufficient capacity to allocate for the advertised capacity of all the thin devices bound to that pool.

- Do not implement space reclamation, available with Enginuity 5874 and later, with pre-allocated or over-subscribed Symmetrix VP (thin) devices that are utilized for host boot/root/swap/dump volumes. Although not recommended, Space reclamation is supported on the listed types of volumes

  Should the customer use space reclamation on this thin device, they need to be aware that this freed space may ultimately be claimed by other thin devices in the same pool and may not be available to that particular thin device in the future.

Cluster configuration

When using high availability in a cluster configuration, it is expected that no single point of failure exists within the cluster configuration and that one single point of failure will not result in data unavailability, data loss, or any significant application becoming unavailable within the cluster. Virtual provisioning devices (thin devices) are supported with cluster configurations; however, over-subscription of virtual devices may constitute a single point of failure if an out-of-space condition should be encountered. To avoid potential single points of failure, appropriate steps should be taken to avoid under-provisioned virtual devices implemented within high availability cluster configurations.

Operating system characteristics

IMPORTANT

The contents of this section also apply to VNX series or CLARiiON systems that support Virtual Provisioning, as well as Symmetrix arrays.

This section describes operating system characteristics when applications write data beyond the provisioned boundary (out-of-space or pool exhaustion).

Most host applications will behave in a similar manner, in comparison to the normal devices, when writing to thin devices. This same behavior can also be observed as long as the thin device written capacity is less than the thin device subscribed capacity. However, issues can arise when the application writes beyond the provisioned boundaries.

With the current behavior of the Operating System AIX 5.2, 5.3, 6.1, JFS, JFS2, VxFS File System, and PowerPath and MPIO Multipathing software, the exhaustion of the thin pool may causes undesired results. Specifics are included below:

- Multi-pathing software
  - **Native MPIO**: When the thin pool reaches 100% capacity, path or paths may become disabled resulting in hung I/O. Server reboot is required in order to correct the condition.
  - **PowerPath**: When the thin pool reaches 100% capacity, path or paths may become disabled resulting in hung I/O. Server reboot is required in order to correct the condition.
  - **VxVM DMP**: No abnormal behavior observed with DMP itself.
- Logical Volume Manager software
  - **Native LVM**: When the thin pool reaches 100% capacity, LVM errors may be observed on the host.
  - **VxVM 5.0 MP1RP3**: When the thin pool reaches 100% capacity, the Volume Manager appears to respond to the Symmetrix as expected by disabling the disk group.
- File System
  - **Native JFS, JFS2**: When the thin pool reaches 100% capacity, file system corruption may occur.
  - **VxFS 5.0 MP1**: VxFS file system corruption has been observed after recovering VxVM from a thin pool that reaches 100% capacity.

Conditions where the host is exposed to pre-provisioned thin devices that had not been bound to the thin pool and presented to the host behave without any problems.

Thin Reclamation

Array-based Thin Reclamation is supported with VMAX 400K, VMAX 200K, VMAX 100K, VMAX 40K, VMAX 20K/VMAX, VMAX 10K (Systems with SN xxx987xxxx), VMAX 10K (Systems with SN xxx959xxxx),VMAXe, and CLARiiON systems and requires Storage Foundation 6.1 and later. VxVM Thin Reclamation functionality is supported with minimum FLARE Release 29, version 04.29.000.5.003 or later. Thin Reclamation functionality is performed on a per-LUN basis.

Host-based Thin Reclamation functionality is supported with VMAX 400K, VMAX 200K, VMAX 100K, VMAX 40K, VMAX 20K/VMAX, VMAX 10K (Systems with SN xxx987xxxx), VMAX 10K (Systems with SN xxx959xxxx),VMAXe, VNX series, and CLARiiON. For VMAX 40K, VMAX 20K/VMAX, VMAX 10K (Systems with SN xxx987xxxx), VMAX 10K (Systems with SN xxx959xxxx), and VMAXe, host-based Thin Reclamation was introduced with EMC Enginuity 5875.135.91.

For IBM operating systems AIX 6.1 and AIX 7.1, host-based Thin Reclamation requires the following software components:

- Symantec Veritas Storage Foundation 6.1 and later
- DMP multipath software
- VMAX 400K with HYPERMAX 5977.250.189 or later
- VMAX 200K with HYPERMAX 5977.250.189 or later
- VMAX 100K with HYPERMAX 5977.250.189 or later
- VMAX 40K with Enginuity 5876.82.57 or later
- VMAX 20K with Enginuity 5876.82.57 or later
- VMAX with Enginuity 5875.135.91 or later
- VMAX 10K (Systems with SN xxx987xxxx) with Enginuity 5876.159.102 or later
- VMAX 10K (Systems with SN xxx959xxxx) with Enginuity 5876.159.102 or later
- VMAXe with Enginuity 5875.198.148e or later
- CX-4 with FLARE 04.30.000.5.509 or later
- VNX with VNX OE for Block v31 or later

Note: Refer to the Dell EMC Simple Support Matrices for the most up-to-date support information

PowerPath is not supported with host-based Thin Reclamation at this time.

PART 2

Midrange Storage Connectivity

Part 3 includes:

CHAPTER 9

IBM AIX Hosts With VNX Series or CLARiiON

This chapter provides information specific to IBM AIX hosts connecting to VNX series or CLARiiON systems.

AIX in a VNX series or CLARiiON environment

This section lists some VNX series and CLARiiON support information specific to the AIX host environment.

Host connectivity

Refer to the Dell EMC Simple Support Matrices or contact your EMC representative for the latest information on qualified hosts and host bus adapters.

Refer to your AIX host documentation for specifications relating to PCI bus limitations and adapter location.

Boot support

Certain AIX servers may utilize VNX series or CLARiiON storage as a boot device. Refer to the Dell EMC Simple Support Matrices or contact your EMC representative for the latest VNX series or CLARiiON storage configurations that support AIX boot.

A boot /root /swap /dump device positioned on VNX or CLARiiON Virtual Provisioning (thin) device(s) is supported. However, some specific processes involving boot /root/swap/dump devices positioned on thin devices should not have exposure to encountering the out-of-space condition.

Host-based processes such as kernel rebuilds, swap, dump, save crash, and Volume Manager configuration operations can all be affected by the dependentthin provisioning out-of-space condition. This exposure is not specific to Dell EMC's implementation of thin provisioning. EMC strongly recommends that the customer avoid encountering the out-of-space condition involving boot / root /swap/dump devices positioned on VNX or CLARiiON VP (thin) devices using the following recommendations

- Virtual Provisioning devices utilized for boot /root/dump/swap volumes must be fully allocated or the VP devices must not be oversubscribed.
- Should the customer use an over-subscribed thin pool, they should understand that they need to take the necessary precautions to ensure that they do not encounter the out-of-space condition.

Storage components

The basic components of a VNX series or CLARiiON storage system configuration are:

- One or more storage systems.
- One or more servers connected to the storage system(s), directly or through hubs or switches.
- For Unisphere/Navisphere 5.X or lower, a Windows NT or Windows 2000 host (called a *management station*) running Unisphere/Navisphere Manager and connected over a LAN to storage system servers and the storage processors (SPs) in VNX series or CLARiiON storage systems. (A management station can also be a storage system server if it is connected to a storage system.)

- For Unisphere/Navisphere 6.X, a host running an operating system that supports the Unisphere/Navisphere Manager browser-based client, connected over a LAN to storage system servers and the SPs in VNX series or CLARiiON storage systems. For a current list of such operating systems, refer to the Unisphere/Navisphere Manager 6.X release notes at Dell EMC Online Support.

Note: The procedures described in this chapter assume that all hardware equipment (such as hubs, switches, and storage systems) used in the documented configuration is already installed and connected.

Required storage system setup

VNX series or CLARiiON system configuration is performed by an EMC Customer Engineer (CE) or the customer through Unisphere/Navisphere Manager. The CE or customer will configure your storage system settings for each SP.

The procedures in this document assume that you have already installed any hubs/switches and storage systems used in this configuration, and connected the storage system SPs to the hub or switch ports.

For a VNX series or CLARiiON storage system in an FC-AL environment, you set the Address ID (AL_PA address) for each SP during storage system setup, as described in the *EMC Navisphere Manager Administrator's Guide.*

Host configuration with IBM native HBAs

VNX series and CLARiiON systems have been qualified with IBM native HBAs. For installation instructions and driver information, consult the appropriate IBM documentation. Refer to the Dell EMC Support Matrices for approved HBAs and configurations.

To establish communication between an AIX server and a VNX series or CLARiiON storage system:

1. Install the HBA(s) and drivers according to the instructions in your IBM documentation.
2. This step depends on the software running in your system:

   Note: CLArray and the VNX series and CLARiiON ODM package cannot coexist on the same system.

   - If the system is running AIX 5.1 or 5.2 with Unisphere/Navisphere 6.5 or earlier with any version of Unisphere/Navisphere, install CLArray software as described under "Utilities" in Chapter 1 of *EMC Storage-System Host Utilities for AIX Administrator's Guide*.

     If the system is already running CLArray and the user is interested in upgrading to the ODM package.

     If the RS/6000 is an SP system, refer to Chapter 2 of *EMC Storage-System Host Utilities for AIX Administrator's Guide*, and follow the steps under either "Installing Utilities on a Shared Filesystem Network" or "Installing Utilities on an Unshared Filesystem Network", depending upon your configuration.
   - If the system is running AIX 5.1 or later with Unisphere/Navisphere revision 6.6.0.1 or later, you must install the VNX series and CLARiiON ODM package.

     Note: For the appropriate version for your configuration, refer to the Dell EMC Support Matrices.

   - The following filesets are distributed within the EMC ODM support package. Always refer to the Dell EMC Support Matrices for the most up-to-date supported solutions.
     - EMC Symmetrix AIX Support Software
       – Standard utility support
     - EMC Symmetrix FCP Support Software
       – IBM Fibre Channel driver support
       – Supports PowerPath
     - EMC Symmetrix FCP MPIO Support Software
       – IBM default PCM MPIO support
     - EMC Symmetrix iSCSI Support Software
       – IBM iSCSI driver support
       – Supports PowerPath
     - EMC Symmetrix HA Concurrent Support
       – Requires HACMP CLVM
     - EMC CLARiiON AIX Support Software

– Standard utility support for VNX series and CLARiiON

– EMC CLARiiON FCP Support Software
  – IBM Fibre Channel driver support for VNX series and CLARiiON
  – Supports PowerPath for VNX series and CLARiiON

– EMC CLARiiON FCP MPIO Support Software
  – IBM default PCM MPIO support for VNX series and CLARiiON

– EMC CLARiiON iSCSI Support Software
  – IBM iSCSI driver support for VNX series and CLARiiON
  – Supports PowerPath for VNX series and CLARiiON

– EMC CLARiiON HA Concurrent Support
  – Requires HACMP CLVM for VNX series and CLARiiON

– EMC Invista AIX Support Software
  – Standard utility support

– EMC Invista FCP Support Software
  – IBM Fibre Channel driver support.
  – Supports PowerPath

– EMC Celerra AIX Support Software
  – Standard utility support

– EMC Celerra FCP Support Software
  – IBM iSCSI driver support

a. Download the appropriate EMC-supplied AIX ODM package version from `ftp.emc.com`:

   1. Issue the command:
      ftp ftp.emc.com
   2. Log in with a username of `anonymous` and your e-mail address as a password.
   3. Issue the command:
      cd /pub/elab/aix/ODM_DEFINITIONS
   4. Issue the command:
      get EMC.AIX.5.x.x.x.tar.Z
   5. Issue the command:
      quit

b. Place EMC Supplied AIX ODM Package Version 5.*x.x.x* into the `/usr/sys/inst.images` directory for installation. Uncompress and untar the package. Enter:

```
# cd /usr/sys/inst.images
# uncompress EMC.AIX.5.x.x.x.tar.Z
# tar -xvf EMC.AIX.5.x.x.x.tar
```

c. Install `EMC.CLARiiON.fcp` using **installp** or **smit**. Enter:

```
# installp -ad /usr/sys/inst.images
      EMC.CLARiiON.fcp
```

3. If you plan on running PowerPath failover support, install the software now, according to the instructions in the *EMCPowerPath for UNIX Installation and Administration Guide*.

4. Install the Unisphere/Navisphere Host Agent and CLI software according to the instructions in the *EMC Navisphere Host Agent and CLI for AIX Version 6.X Installation Guide*.
5. Connect the HBA(s) to the CLARiiON storage system according to your planned configuration. If the devices are being attached via a switch, configure the zone sets on the switch accordingly.
6. Check for the current `arraycommpath` setting on the array with the following command:

   navicli -h *<hostname_or_IP_address_of_SP_A>* **arraycommpath**

   If `arraycommpath` is disabled (return value of 0), then enable it with the following command:

   navicli -h *<hostname_or_IP_address_of_SP_A>* **arraycommpath 1**
7. From the host, run the **cfgmgr** command.
8. Run the command:
   lsdev -Cc disk

   There should be a LUNZ device for each SP port that the host HBAs are connected to. Each LUNZ device will have a corresponding hdisk device name. Remove these with the following command:

   rmdev -Rdl <*hdisk_name***>**
9. Manage the array via Unisphere/Navisphere Manager. Right-click the array icon and select **Connectivity Status** from the pull-down menu.

   You should see entries for each of your HBAs, showing that they are present on the fibre, but are not registered.
10. To register each HBA, go to the host and recycle the Unisphere/Navisphere Host Agent with this command sequence:

    /etc/rc.agent stop
    /etc/rc.agent start
11. Return to Unisphere/Navisphere Manager, right-click the array, and select **update now**.

    This refreshes the information in your browser window. Re-select the **Connectivity Status** option and verify that the each HBA entry is now registered and has the correct hostname associated with it.
12. Click the **Group Edit** option in the **Connectivity Status** window.

    Within the **Group Edit Initiators** pop-up:

    a. Select the HBAs from your host in the **Available** list and move them to the **Selected** list.

    b. Under **New Initiator Information**, set **Arraycommpath** to **enabled**.

c. Failovermode should be set according to your configurations:

- Set to **1** for PowerPath non-NDU configurations.
- Set to **1** for MPIO configurations, NDU configurations.
- Set to **1** for Veritas DMP configurations (requires VxVM 5.0.1.3 and later).
- Set to **2** for Veritas DMP configurations requires VxVM 5.0.1.2 and earlier).
- Set to **3** for PowerPath, NDU configurations.
- Set to **4** for PowerPath, MPIO ALUA configurations, NDU configurations.

Note: For more information on CLARiiON NDU with PowerPath, refer to Chapter 10, "VNX Series or CLARiiON Nondisruptive Upgrade (NDU)."

d. Under **These HBAs belong to**, select **Existing Host** and make sure that your hostname is displayed. Then click **OK**.

13. Bind LUNs according to your planned storage configuration. Refer to the *EMC Navisphere Manager Version 6.X Administrator's Guide* for more information on this step.
14. If EMC Access Logix™ is enabled for your configuration, create a storage group for this host and assign the LUNs to it.

    Once LUNs are assigned to the host, LUNZ devices will no longer be created after the **cfgmgr** command is run as long as a LUN0 is present in the storage group.

What next? Proceed to "Making LUNs available to AIX" on page 260.

Making LUNs available to AIX

This section explains how to make newly bound storage system LUNs available to AIX. It assumes that you have successfully completed the steps under "Host configuration with IBM native HBAs" on page 256.

AIX treats a bound LUN as a single disk. As you work with AIX, you can treat any bound LUN just as you would any newly acquired disk.

Verifying disk configurations

A disk is configured and made available to AIX through the device configuration method included with CLArray. The configuration method defines and configures all LUNs that are assigned to the host by the VNX series or CLARiiON storage system as AIX *hdisk* devices.

To configure previously bound LUNs under AIX, run the following command:

/usr/lpp/EMC/CLARiiON/bin/emc_cfgmgr

When this is complete, run the following command:

lsdev -Cc disk

AIX will display a list of disk entries, each including name, status, location, and description. An internal SCSI disk drive would appear similar to the following:

```
hdisk0      Available      40-60-00-4,0         16 Bit SCSI Disk Drive
```

If your system is running CLArray, a CLARiiON RAID 5 LUN would appear similar to this:

```
hdisk2      Available      11-08-01       FC CLArray RAID 5 Disk Group
```

If your system is running the CLARiiON ODM package, the same RAID 5 LUN would appear similar to this:

```
hdisk2      Available      11-08-01       EMC CLARiiON FCP RAID 5 Disk
```

Note: `11-08-01` in the above examples is the location code. Location codes are in the format:

`AB-CD-EF`
where:

`AB` identifies which bus into which the adapter is plugged.

`CD` identifies which slot in the bus into which the adapter is plugged.

`EF` is the pseudo qualifier which is specified by the high-order FC-AL Address ID switch on the storage system.

For complete information on location codes, refer to your AIX documentation.

An hdisk device listed as `Available` is in a normal state and ready to accept I/O. If an hdisk is listed as `defined`, it is not currently available to AIX and is not ready for I/O requests.

You can obtain the same information through `SMIT` by entering the command **smit disk** at the shell prompt and selecting **List All Defined Disks** from the **Fixed Disk** menu. After viewing the status screen that `SMIT` displays, press **F3** or **ESC-0** (zero) to exit the display.

PowerPath devices

If more than one path exists for a given LUN, AIX creates a unique hdisk for each path. If this is the case with PowerPath installed, an hdiskpower device is created for each set of hdisks that map to the same physical LUN, in order to handle failover functions for this VNX series or CLARiiON device. The hdiskpower devices also appear along with the hdisks when running the **lsdev -Cc disk** command.

If PowerPath is installed and multiple paths exist between the host and VNX series or CLARiiON storage, but hdiskpower devices are not present after running the **emc_cfgmgr** command, run the **powermt config** command to configure them.

Consult the *PowerPath for UNIX Installation and Administration Guide* for more information on PowerPath.

Creating volume groups, logical volumes, and filesystems

After VNX series or CLARiiON devices are fully configured and available to AIX, they can be placed into volume groups and have logical volumes and filesystems created on them for storing customer data.

Refer to "Configuring disks using SMIT" on page 199 for further information on creating volume groups, logical volumes. and filesystems.

LUN expansion

Unisphere/Navisphere offers two methods for expanding VNX series or CLARiiON LUN capacity: RAID Group and MetaLUN expansion. For existing AIX filesystems to recognize and take advantage of any additional capacity provided by VNX series or CLARiiON LUN expansion, AIX servers must have AIX Revision 5.2 ML-01 APAR IY1957 or later installed.

The steps for performing this expansion are:

1. Perform LUN expansion (RAID Group or MetaLUN) as detailed in the *EMC Navisphere Manager 6.X Administrator's Guide*.
2. From AIX, unmount the filesystem and vary off the volume group (using the **umount** and **varyoffvg** commands).
3. From AIX, run the **emc_cfgmgr** command.
4. Run the **inq** command to verify that AIX recognizes the hdisk's increased capacity.
5. Vary on the volume group with the **varyonvg** command.
6. Run the **chvg -g** <*vg_name*> command so the volume group's new capacity is recognized.
7. Run the **lslv** <*lv_name*> command and note the PP SIZE and PPs. Multiply the two in order to approximate the new size of the volume in megabytes. Convert this number to 512-byte blocks by multiplying by 2048. (If your logical volume PP SIZE is less than 1 MB, then adjust the calculation accordingly.)
8. Mount the filesystem and run the **chfs -a size=**<*new_size_from_step_7*> <*fs_name*> command. This will expand the filesystem, using all new space available on the device.

Note: MetaLUN expansion is the recommended method for increasing LUN capacity. RAID Group expansion supports only the expansion of LUN capacity for RAID Groups that contain a single LUN.

Fast I/O failure for Fibre Channel devices

As of AIX release 5.2 Maintenance level 2 (APAR IY48488), IBM has introduced a new feature called *Fast I/O Failure*. If the Fibre Channel adapter driver detects a link event, such as a lost link between a storage device and a switch, the driver waits approximately 15 seconds to allow the fabric to stabilize. After this wait period, the driver detects that the device is not on the fabric and begins failing all I/Os. Any new I/Os are failed immediately by the adapter until the driver detects that the device has returned.

Fast I/O Failure is controlled by the new attribute **fc_err_recov** in the fscsi that is defined in the ODM. The default setting of this attribute is `delayed_fail`, which in effect disables this feature. Setting the attribute to `fast_fail` enables the Fast Failover feature.

Here is an example of enabling the feature on the fscsi0 adapter:

```
chdev -l fscsi0 -a fc_err_recov=fast_fail -P
```

The above command would have to be repeated for all present fscsi adapters in the system.

The Fast Fail feature is now enabled for the fscsi0 adapter. When the adapter driver receives an event from the switch that there has been a link event (RSCN), the driver invokes the Fast Fail logic.

Fast Fail functionality is desirable when multipathing software (such as EMC PowerPath) is installed. Setting the `fc_err_recov` attribute to `fast_fail` should decrease the failure times due to a link failures.

In a single direct-connect path environment, it is recommended that the Fast Fail feature be disabled.

Requirements

The requirements for Fast I/O Failure support are:

- Fast I/O Failure is supported only in a switched environment. It is not supported in arbitrated loop environments.
- FC 6227 adapter firmware — level 3.22A1 or greater.
- FC 6228 adapter firmware — level 3.82a1 or greater.
- FC 6239 adapter firmware — level 1.00X5 or greater.
- FC 5716 adapter firmware — level 1.90A4 or greater.
- FC 1977 adapter firmware — level 1.90A4 or greater.
- Newer released HBAs contain base support.

Failure to meet these requirements can cause the fscsi device to log an error log of type `INFO` indicating that one of these requirements was not met and that fast fail *has not* been enabled.

Dynamic Tracking of Fibre Channel devices

As of AIX release 5.2 Maintenance level 2 (APAR IY48488), IBM has introduced a new feature called *Dynamic Tracking of Fibre Channel Devices*. Previous versions of the operating system required an administrator to delete hdisks and adapters before making changes to the fabric that would result in a different N_Port ID. For example, if moving a fibre cable from one switch port to another, a device reconfiguration would be required.

If Dynamic Tracking is enabled, the Fibre Channel adapter driver will detect the occurrence of a change in the N_Port ID and automatically reroute I/Os for that device to the new address. Examples of events that can cause an N_Port ID to change are moving a cable between a switch and storage device from one switch port to another, connecting two separate switches via an interswitch link (ISL), and possibly rebooting a switch.

The Dynamic Tracking for Fibre Channel devices is controlled by a new fscsi attribute called `dyntrk`. The default setting for this attribute is **no** (disabled), Setting this attribute to **yes** enables the feature for the named adapter.

The following example enables Dynamic Tracking for adapter fscsi0

```
chdev -l fscsi0 -a dyntrk=yes -P
```

Requirements

Requirements for Dynamic Tracking support are:

- Fast I/O Failure is supported only in a switched environment. It is not supported in arbitrated loop environments.
- FC 6227 adapter firmware — level 3.22A1 or greater.
- FC 6228 adapter firmware — level 3.82a1 or greater.
- FC 6239 adapter firmware — level 1.00X5 or greater.
- FC 5716 adapter firmware — level 1.90A4 or greater.
- FC 1977 adapter firmware — level 1.90A4 or greater.
- Newer released HBAs contain base support.
- The device World Wide Name (Port Name) and Node Name must remain constant, and the World Wide Name must be unique. Changing the World Wide Name or Node Name of an available or online device could result in I/O failures.
- Each Fibre Channel storage device instance must have `world_wide_name` and `node_name` attributes, both of which are included in EMC AIX ODM package 5.1.0.0 and later.

  Updated filesets that contain the `sn_location` attribute (discussed in the list item) should also be updated to contain the `world_wide_name` and `node_name`.

- The storage device must provide a reliable method to extract a unique serial number for each LUN. The AIX Fibre Channel device drivers will not autodidact serial number location, so the method for serial number extraction must be explicitly provided by any storage vendor in order to support dynamic tracking for their devices.

  This information is conveyed to the drivers using the `sn_location` ODM attribute for each storage device. If the disk or tape driver detects that the `sn_location` ODM attribute is missing, an error log of type `INFO` will be generated and dynamic tracking *will not* be enabled.

  Note: The `sn_location` attribute may be non-displayable, so running the **`lsattr`** command on an hdisk, for example, may not show the attribute, but it may, indeed, be present in ODM.

- The Fibre Channel device drivers will be able to track devices on a SAN fabric (a fabric as seen from a single host bus adapter) if the N_Port IDs on the fabric stabilize within about 15 seconds. If cables are not reseated or N_Port IDs continue to change after the initial 15 seconds, I/O failures could result.
- Devices will not be tracked across host bus adapters. Devices will only track if they remain visible from the same HBA that they were originally connected to.

  For example, if device A were moved from one location to another on fabric A attached to host bus adapter A (i.e., its N_Port on fabric A changes), the device would seamlessly be tracked without any user intervention and I/O to this device can continue.

  However, if a device A is visible from HBA A but not from HBA B, and device A is moved from the fabric attached to HBA A to the fabric attached to HBA B, device A will not be accessible on fabric A nor on fabric B. User intervention would be required to make it available on fabric B by invoking cfgmgr. The AIX device instance on fabric A would no longer be usable, and a new device instance on fabric B would be created. This device would have to manually be added to volume groups, multipath device instances, etc. In essence, this is the same as removing a device from fabric A and adding a new device to fabric B.
- No dynamic tracking will performed for Fibre Channel dump devices while an AIX system dump is in progress. In addition, dynamic tracking is not supported during boot or during cfgmgr invocations. SAN changes should not be made while any of these operations are in progress.
- Once devices are tracked, ODM will potentially contain stale information as SCSI IDs in ODM will no longer reflect actual SCSI IDs on the SAN. ODM will remain in this state until cfgmgr is run manually or the system is rebooted, provided all drivers, including any third-party Fibre Channel SCSI target drivers, are dynamic-tracking capable. If cfgmgr is run manually, cfgmgr must be invoked on all affected fscsi devices, which can easily be accomplished by running **`cfgmgr`** without any options, or by invoking cfgmgr on each fscsi device individually.

  Note: Running **`cfgmgr`** at run time to recalibrate the SCSI IDs may not update the SCSI ID in ODM for a storage device if the storage device is currently opened, such as when volume groups are varied on. Run `cfgmgr` on devices that are not opened or the system should be rebooted to recalibrate the SCSI IDs. Note that stale SCSI IDs in ODM have no adverse affect on the FC drivers and recalibration of SCSI IDs in ODM is not necessary for

the FC drivers to function properly. Any applications that communicate with the adapter driver directly via ioctl calls and use the SCSI ID values from ODM, however, need to be updated as indicated in the next bullet to avoid using potentially stale SCSI IDs.

- Even with dynamic tracking enabled users are strongly encouraged to make SAN changes, such as cable moves/swaps and establishing ISL links, during maintenance windows. Making SAN changes during full production runs is discouraged. This is due to the fact that there is a short interval of time to perform any SAN changes. Cables that are not reseated correctly, for example, could result in I/O failures. Performing these operations during a time of little/no traffic minimizes impact of I/O failures due to misplugging of cables, taking too long to re-cable, etc.

Fast Fail and Dynamic Tracking interaction

Although Fast Fail and Dynamic Tracking of Fibre Channel devices are technically separate features, the enabling of one may change the interpretation of the other in certain situations. The following table shows the behavior exhibited by the Fibre Channel drivers with the various permutations of these settings.

dyntrk	fc_err_recov	Fibre Channel Driver Behavior
no	delayed_fail	This is the default setting. This is legacy behavior existing in previous versions of AIX. The FC drivers will not recover if the scsi_id of a device changes, and I/Os will take longer to fail when a link loss occurs between a remote storage port and switch. This may be desirable in single-path situations if dynamic tracking support is not a requirement.
no	fast_fail	If the driver receives a Registered State Change Notification (RSCN) from the switch, this could indicate a link loss between a remote storage port and switch. After an initial 15 second delay, the FC drivers will query to see if the device is on the fabric. If not, I/Os will be flushed back by the adapter. Future retries or new I/Os will fail immediately if the device is still not on the fabric. If the FC drivers detects the device is on the fabric but the scsi_id has changed, the FC device drivers will not recover, that is, I/Os will be failed with PERM errors.
yes	delayed_fail	If the driver receives a Registered State Change Notification (RSCN) from the switch, this could indicate a link loss between a remote storage port and switch. After an initial 15 second delay, the FC drivers will query to see if the device is on the fabric. If not, I/Os will be flushed back by the adapter. Future retries or new I/Os will fail immediately if the device is still not on the fabric, although the storage driver (disk, tape) drivers may inject a small delay (2-5 seconds) between I/O retries. If the FC drivers detects the device is on the fabric but the `scsi_id` has changed, the FC device drivers will reroute traffic to the new scsi_id.

dyntrk	fc_err_recov	Fibre Channel Driver Behavior (continued)
yes	fast_fail	If the driver receives a Registered State Change Notification (RSCN) from the switch, this could indicate a link loss between a remote storage port and switch. After an initial 15 second delay, the FC drivers will query to see if the device is on the fabric. If not, I/Os will be flushed back by the adapter. Future retries or new I/Os will fail immediately if the device is still not on the fabric. The storage driver (disk, tape) will likely not delay between retries. If the FC drivers detects the device is on the fabric but the scsi_id has changed, the FC device drivers will reroute traffic to the new `scsi_id`.

Note that with dynamic tracking disabled, there is a marked difference between the **`delayed_fail`** and **`fast_fail`** settings of the **`fc_err_recov`** attribute. However, with dynamic tracking enabled, the setting of the **`fc_err_recov`** attribute is less significant. This is because there is some overlap in the dynamic tracking and fast fail error recovery policies. As such, enabling dynamic tracking inherently enables some of the fast fail logic.

The general error recovery procedure when a device is no longer reachable on the fabric is the same for both **`fc_err_recov`** settings with dynamic tracking enabled, with the minor difference that the storage drivers may choose to inject delays between I/O retries if **`fc_err_recov`** is set to **`delayed_fail`**. This will increase the I/O failure time by some additional amount depending on the delay value and number of retries before permanently failing the I/O. With high I/O traffic, however, the difference between **`delayed_fail`** and **`fast_fail`** may be more noticeable.

SAN administrators may want to experiment with these settings to find the correct combination of settings for their environment.

CHAPTER 10

VNX Series or CLARiiON Nondisruptive Upgrade (NDU)

This chapter contains the following information on the VNX series or CLARiiON nondisruptive upgrade.

Introduction

Interactions between FLARE, PowerPath, and AIX that occur during the nondisruptive upgrade (NDU) process previously led to a perceived loss of connectivity which periodically resulted in AIX and/or host applications reporting errors during online upgrades. It must be noted that the actual software upgrade to the VNX series or CLARiiON system did not fail. The software upgrade worked successfully but AIX or host applications, or both, may have been impacted.

EMC, in cooperation with IBM, has designed a solution for these interaction problems. A customer must perform at least one more OFFLINE upgrade of FLARE, PowerPath, and AIX drivers, and change their settings to enable future array software upgrades to be performed while the AIX servers are ONLINE.

PowerPath 4.5.1 and PowerPath SE 4.5.1 and later support non-disruptive upgrades (NDU) of CLARiiON storage system software. PowerPath 5.3 and later support non-disruptive upgrades (NDU) of VNX Series storage system software.

Note: For the most up to date guidelines, restrictions, and approved configurations for NDU, refer to the Dell EMC Support Matrices.

Note: Very rarely an AIX host whose boot disk and paging space reside on a VNX series or CLARiiON system may experience a hang during a non-disruptive Upgrade.

In EMC testing, the frequency of AIX hangs under heavy I/O load was less than 1 in 1000 non-disruptive upgrades. Reducing I/O load for the AIX host before a non-disruptive upgrade reduces the chance of the hang occurring. If a host hang occurs, it is necessary to reboot the host to restore array access.

Non-disruptive upgrades in a VNX series environment with PowerPath

This section includes the NDU checklist and information for preparing the AIX host for non-disruptive upgrades in a VNX series environment.

Configurations running MPIO or Veritas Volume Manager (VxVM) without PowerPath are exempt from the following settings outlined in this chapter.

AIX servers running PowerPath and VxVM must adhere to the settings defined in this section.

NDU checklist

If your system meets the following minimum requirements listed, then you are already positioned to perform non-disruptive upgrades in a VNX series environment.

If you do not meet all of these requirements, refer to the section, "AIX host preparation" on page 273 for instructions on how to configure these settings.

Minimum requirements

- Array software
  - VNX OE for Block v31or later
- For 6227 HBAs num_cmd_elems = 1024
- For 6228 HBAs num_cmd_elems = 2048
- For 6227 and 6228 max_xfer_size = 0x1000000
- For all HBAs fc_err_recov = fast_fail

IMPORTANT

HBAs released after 6227/6228 support the default settings for num_cmd_elems and max_xfer_size.

- Failovermode 4 for initiator ports for all storage groups on the host
- Base AIX versions as listed in the *EMC Support Matrix* (ESM). No additional APARs required.

Non-disruptive upgrades in a CLARiiON environment with PowerPath

This section includes the NDU checklist and information for preparing the AIX host for non-disruptive upgrades in a CLARiiON environment with PowerPath.

Configurations running MPIO or Veritas Volume Manager (VxVM) without PowerPath are exempt from the following settings outlined in this chapter.

AIX servers running PowerPath and 6227 and 6228 HBAs must adhere to the settings defined in this section.

NDU checklist

If your system meets the following minimum requirements listed, then you are already positioned to perform non-disruptive upgrades in a CLARiiON environment.

If you do not meet all of these requirements, refer to the next section, section, "AIX host preparation" on page 273 for instructions on how to configure these settings.

Configurations running MPIO or Veritas Volume Manager (VxVM) without PowerPath are exempt from the following settings outlined in this chapter.

NDU for configurations running MPIO is only supported for AIX 5.3 and later.

AIX servers running PowerPath and VxVM must adhere to the settings defined in this section.

Refer to the Dell EMC Support Matrices for specific versions of VxVM supported for NDU.

Minimum requirements

- Array software
  - 2.19.xxx.5.016 or later
- For 6227 HBAs num_cmd_elems = 1024
- For 6228 HBAs num_cmd_elems = 2048
- For 6227 and 6228 max_xfer_size = 0x1000000
- For all HBAs fc_err_recov = fast_fail

IMPORTANT

HBAs released after 6227/6228 support the default settings for num_cmd_elems and max_xfer_size.

- Failovermode 3 for initiator ports for all storage groups on the host
- AIX 5.1
  - ML 8 or later with APARs IY73975, IY73977, IY75022, and IY73978, EMC AIX ODM kit 5.1.0.3 or later
- AIX 5.2
  - ML 5 or later with APARs IY71263, IY71265, IY73709, and IY73782, EMC AIX ODM kit 5.2.0.3 or later
- AIX 5.3
  - ML 2 or later with APARs IY71303, IY71653, IY73637, EMC AIX ODM kit 5.2.0.3 or later
- AIX PowerPath 4.5.1 or AIX PowerPath SE 4.5.1 (included in CLARiiON AIX utility kit) or later
- Less than 16 paths maximum to the array per AIX host
- Less than 100 LUNs maximum per AIX host for AIX 5.1 and 5.2

AIX host preparation

It is assumed that any AIX host system software will be placed in the /usr/sys/inst.images directory for installation.

IMPORTANT

This procedure is a disruptive process and host I/O access to the array needs to be restricted.

Once you make the software and Fibre Channel device changes described in this section, the AIX host system will be ready for future online upgrades of CLARiiON storage system software. If the AIX Fibre Channel devices are removed from the configuration database, you would need to reapply changes to the device characteristics.

To prepare an AIX host for NDU:

1. Login as root.
2. Unmount any file systems and vary off any volume groups associated with the storage array.

   Note: Host applications should also be offline.

3. Remove all PowerPath devices. Enter:

   ```
   powermt remove dev=all
   # lsdev -Ctpower -cdisk -F name | xargs -n1 rmdev -dl
   savebase -v
   ```

4. Remove all CLARiiON hdisks, and repeat this step for all hdisks on the CLARiiON system. Enter:

   ```
   rmdev -dl hdiskX
   ```

5. Place all IBM Fibre Channel HBAs into a defined state, and repeat this step for all HBAs on the host. Enter:

```
rmdev -l fcsX –R
```

6. Change fscsi and fcs attributes for each IBM Fibre Channel adapter, and repeat this step for all adapters on the host. (Settings for fast_fail apply only to Fibre Channel switch environments.) Enter the following commands:

   ```
   chdev -l fcsX -a num_cmd_elems=2048
   chdev -1 fcsX -a max_xfer_size=0x1000000
   ```

   Note: For FC6227 HBAs, the maximum value for num_cmd_elems is 1024.

   ```
   chdev -l fscsiX -a fc_err_recov=fast_fail
   ```

7. If the CLARiiON storage system is not at version 02.19.xxx.5.016 or later, update the CLARiiON FLARE version to 02.19.*XXX*.5.016 or later.

8. Remove any existing version of Navisphere Agent from the AIX host server.

   ```
   installp -u NAVIAGENT NAVICLI
   ```

9. Install Navisphere Agent version 6.19.0.4 on the AIX host server. Enter:

   ```
   installp -agXd /usr/sys/inst.images NAVIAGENT NAVICLI
   ```

10. Using Navisphere CLI, set the failovermode value to 3 on the storage array for the AIX host system. Enter:

    ```
    navicli -h <arrayIPaddress> storagegroup –sethost -host <hostname>
    -failovermode 3 –o
    ```

11. Using the following example install the required IBM APARs according to IBM instructions. This needs to be done for each individual fix.

    For example, for AIX 5.1 fixes (IY73975, IY73977, IY75022, and IY73978), enter:

    ```
    instfix –k fix_name –d location_of_fix
    instfix -k IY73975-d /usr/sys/inst.images
    ```

    - For AIX 5.2 fixes substitute **(IY71263, IY71265, IY73709, and IY73782)**
    - For AIX 5.3 fixes substitute **(IY71303, IY71653, and IY73637)**

12. If the existing EMC.ODM software is not at the revision stated in "NDU checklist" on page 271, remove the current ODM and install the appropriate revision. Enter:

    ```
    installp -u EMC.CLARiiON.fcp
    installp -agXd /usr/sys/inst.images EMC.CLARiiON.fcp
    ```

13. Install PowerPath 4.5.1 or later Enter:

    ```
    installp -agXd /usr/sys/inst.images EMCpower
    ```

    For complete installation instructions, refer to the *EMC PowerPath for AIX Installation and Administration Guide*.

14. Reboot the AIX host system.

15. Verify the configuration. Enter this series of commands:

    ```
    lsdev -Cc disk
    powermt display
    lsattr -El fscsiX
    lsattr -El fcsX
    ```

```
lslpp -L EMC.CLARiiON.fcp.rte
lslpp -L EMCpower.base
oslevel –r
```

16. Using the following example verify the installation of APARs, this command must be run for each individual fix.

    - For AIX 5.1, ensure these APARs: **IY73975, IY73977, IY75022, and IY73978** are installed:

      ```
      instfix -ik fix_name
      instfix -ik IY73975
      ```

    - For AIX 5.2, ensure these APARs: **IY71263, IY71265, IY73709, and IY73782** are installed.

    - For AIX 5.3, ensure these APARs: **IY71303, IY71653, and IY73637**are installed**.**

17. Make sure the num_cmd_elems values of FC adapters are set to maximum. Enter:

    ```
    lsattr -a num_cmd_elems -El fcsX
    lsattr -a max_xfer_size -El fcsX
    ```

    Note: For FC6227 HBAs, the maximum value for num_cmd_elems is 1024.

    Output:

    ```
    num_cmd_elems 2048 Maximum number of COMMANDS to queue to the
    adapter True
    ```

    ```
    max_xfer_size 0x1000000 Maximum Transfer Size True
    ```

18. In Fibre Channel switch environments only, ensure fast fail is set for all HBAs. Enter:

    ```
    lsattr –a fc_err_recov -El fscsi1
    ```

    Output:

    ```
    fc_err_recov fast_fail FC Fabric Event Error RECOVERY Policy True
    ```

19. Check the CLARiiON port failovermode. Enter:

    ```
    navicli -h <IP> port -list -hba -failovermode
    ```

The failovermode should be set to **3** for each of the HBAs attached to the AIX host.

CHAPTER 11

AIX with Unity Support

This chapter is a reference for Dell EMC Unity array support in the IBM AIX environment.

Overview

Dell EMC Unity™ Hybrid and All Flash storage systems implement an integrated architecture for block, file, and VMware virtual volumes (VVols) with concurrent support for native NAS, iSCSI, and Fibre Channel protocols based on the powerful new family of Intel E5-2600 processors. Each system leverages dual storage processors, full 12 Gb SAS back-end connectivity, and an EMC multi-core operating environment to deliver unparalleled performance and efficiency. Additional storage capacity is added via Disk Array Enclosures (DAEs).

Refer to the following documents on Dell EMC Online Support and Dell EMC E-Lab Navigator for configuration and administration operations:

- *Unity Hybrid and Unity All Flash Installation Guide*
- Unity Family Unisphere CLI User Guide
- White Paper: Unity: Unisphere Overview
- White Paper: Unity: Best Practices Guide
- White Paper: Unity: Introduction to the Unity Platform
- *EMC Unity 600F/600/500F/500/400F/400/300F/300/VSA Dell EMC Simple Support Matrix.*

Prerequisites

Complete the following prerequisites before configuring Unity in the IBM AIX environment.

Host connectivity

Refer to the Dell EMC Simple Support Matrices or contact your Dell EMC representative for the latest information on qualified hosts, host bus adapters, and connectivity equipment.

- Supported AIX version: AIX 6.1, AIX7.1, and AIX 7.2.
- AIX 5.3 TL12 can be conditionally supported. AIX 5.3 is now End-of-Server-Life, and IBM no longer provides extended support agreements for this older OS version. If you encounter significant issues with AIX 5.3, upgrade to a newer supported AIX version.

Note: For other technique levels of AIX 5.3, the configuration with Unity requires RPQ approval.

ODM version requirement

PowerPath and Veritas DMP can be supported with a Unity connection to AIX hosts. This requires a minimum Dell EMC ODM 5.3.0.5 or 6.0.0.0.

- Use ODM 5.3.0.5 or later for AIX 5.3, AIX 6.1, and VIOS.
- Use ODM 6.0.0.0 or later for AIX 7.1 and AIX 7.2.

FC connection configuration

Install the Unity ODM file sets to configure the AIX host to use Unity disks over Fibre Channel. Also install different filesets for PowerPath/DMP and the MPIO environment. Refer to the following steps for Unity FC configuration.

Note: Unity presents LUNs to the host with the **CLARiiON** device type, and the description of a disk is **EMC CLARiiON disk**.

Install the ODM kit

1. Remove all of the Unity disks, if they are recognized as an **Other FC SCSI Disk Drive** by the system:

   ```
   lsdev -Cc disk | grep "Other FC SCSI Disk Drive" | awk {'print $1'}
   | xargs -n1 rmdev -dl
   ```

2. Uninstall the old **CLARiiON ODM** filesets by typing the following command:

   ```
   installp -u EMC.CLARiiON.*
   ```

3. Download the AIX ODM package version 5.3.*x.x* or 6.0.*x.x* from the ftp server at **ftp.emc.com**.

   e. Type the following command:

   ```
   ftp ftp.emc.com
   ```

   f. Log in with an **anonymous** username and your email address as a password.

   g. Type the following command:

   ```
   cd /pub/elab/aix/ODM_DEFINITIONS
   ```

   h. Type the following command:

   ```
   get EMC.AIX.5.3.x.x.tar.Z
   ```

   i. Type **quit**.

4. Insert AIX ODM package version 5.3.*x.x* or 6.0.*x.x* into the `/usr/sys/inst.images` directory for installation. Uncompress and untar the package, and then type the following command:

   ```
   cd /usr/sys/inst.images
   uncompress EMC.AIX.5.3.x.x.tar.Z
   tar -xvf EMC.AIX.5.3.x.x.tar
   ```

5. Type the following command to Install the desired Unity/CLARiiON ODM filesets:

   ```
   rm .toc
   inutoc .
   ```

   - Type the following if using PowerPath/DMP:

     ```
     installp -ac -gX -d . EMC.CLARiiON.aix.rte
     ```

     ```
     installp -ac -gX -d . EMC.CLARiiON.fcp.rte
     ```

- Type the following if using MPIO:

  ```
  installp -ac -gX -d . EMC.CLARiiON.aix.rte
  installp -ac -gX -d . EMC.CLARiiON.fcp.MPIO.rte
  ```

  Note: You can also use **smitty** for the filesets installation.

Scan and verify Unity LUNs

Use the following commands:

- Type the **cfgmgr** AIX system command to scan disks.
- Type the **lsdev -Cc disk** command to verify that all FC disks are configured properly and show the recognized Unity disks.
- If using PowerPath/DMP, the result will be as follows:

  ```
  hdisk1      Available          EMC CLARiiON FCP VRAID Disk
  hdisk2      Available          EMC CLARiiON FCP VRAID Disk
  ```

- If using MPIO, the result will be as follows:

  ```
  hdisk1      Available          EMC CLARiiON FCP MPIO VRAID Disk
  hdisk2      Available          EMC CLARiiON FCP MPIO VRAID Disk
  ```

iSCSI connection configuration

Install the Unity ODM filesets to configure the AIX host to use Unity iSCSI disks. Refer to following steps for Unity iSCSI configuration.

PowerPath and DMP, but not MPIO, are supported for multipath software.

Refer to the IBM support page for SCSI software initiator configuration, and Unity documents for iSCSI interface configuration.

Install the EMC ODM kit

1. Type the following command to remove all the Unity iSCSI disks, if they're recognized as **Other iSCSI Disk Drive** by the system:

   ```
   lsdev -Cc disk | grep "Other iSCSI Disk Drive" | awk {'print $1'} |
   xargs -n1 rmdev -dl
   ```

2. Uninstall the old CLARiiON ODM filesets by typing the following command:

   ```
   installp -u EMC.CLARiiON.*
   ```

3. Download the AIX ODM package version 5.3.*x.x* or 6.0.*x.x* from the ftp server at **ftp.emc.com**.

   a. Type the following command:

      ```
      ftp ftp.emc.com
      ```

   b. Log in with an **anonymous** username and your email address as a password.

   c. Type the following command:

      ```
      cd /pub/elab/aix/ODM_DEFINITIONS
      ```

   d. Type the following command:

      ```
      get EMC.AIX.5.3.x.x.tar.Z
      ```

   e. Type **quit**.

4. Insert AIX ODM package version 5.3.*x.x* or 6.0.*x.x* into the `/usr/sys/inst.images` directory for installation. Uncompress and untar the package, and then type the following command:

   ```
   cd /usr/sys/inst.images
   uncompress EMC.AIX.5.3.x.x.tar.Z
   tar -xvf EMC.AIX.5.3.x.x.tar
   ```

5. Type the following command to Install the desired EMC Unity/CLARiiON ODM filesets:

   ```
   rm .toc
   inutoc .
   ```

   - Type the following command to install PowerPath/DMP, which is the only supported software:

     ```
     installp -ac -gX -d . EMC.CLARiiON.aix.rte
     ```

     ```
     installp -ac -gX -d . EMC.CLARiiON.fcp.rte
     ```

- Install the supported PowerPath version after the Unity iSCSI filesets are installed successfully.

Note: You can also use **`smitty`** for the filesets installation.

Scan and verify Unity LUNs

1. Type the AIX system command to scan iSCSI disks:

   ```
   cfgmgr -l iscsi0
   ```

2. Type the **`lsdev -Cc disk`** command.

   The following verifies that all iSCSI disks are configured properly, with the recognized Unity iSCSI disks:

   ```
   hdisk1      Available          EMC CLARiiON iSCSI VRAID Disk
   hdisk2      Available          EMC CLARiiON iSCSI VRAID Disk
   ```

Booting the host from a Unity disk

This section covers the requirements and operating steps for booting the host from a Unity disk.

Requirements

The following information provides general guidelines, requirements, and instructions for booting AIX from the Unity LUN. Refer to the Dell EMC Support Matrix for supported hardware.

1. If the host stuck at IBM reference code 0554 and does not boot, try to configure only one logical path from the Unity LUN to the host.
2. Unity arrays require Persistent PID Binding. The Brocade switch must enable the WWN-based persistent PID feature; otherwise the boot from Unity array will fail.
3. Booting from a SAN is supported for Unity arrays when a FC connection is configured; An iSCSI connection is not supported.

Setting up booting from a SAN

1. Check the sub-paths of the disk, if PowerPath/DMP is used in your system. Record all the sub-paths.
   - Type the following command for a PowerPath installed system:

     `powermt display dev=hdiskpower#`
   - Type the following command for a Veritas DMP installed system:

     `vxdmpadm getsubpaths dmpnodename=clariion#`
2. Release the sub-paths from the control of PowerPath/DMP.
   - Type the following command for a PowerPath installed system:

     `powermt remove dev=hdiskpower#`

     `rmdev -dl hdiskpower#`
   - Type the following command for a Veritas DMP installed system:

     `vxdiskunsetup -C clariion#`

     `vxdisk rm clarion#`
3. Type the following command to clone the system into an available sub-path of the multipath disk:

   `alt_disk_install -C hdisk#`

   Note: We recommend configuring only one sub path.
4. Type the **`lspv | grep rootvg`** command to check if an alternate installation disk was cloned successfully.

   Note: There should be an **`altinst_rootvg`** volume group.

This is the result of running the command:

```
hdisk0          00c548453a4d0f10        rootvg          active
hdisk1          00c548458c4a1ab5        altinst_rootvg
```

5. Type the **bootlist -om normal** command to check the boot list.

   This shows the list that the system that has been cloned to:

   ```
   hdisk1
   ```

6. Reboot your system.
7. Type the following command to add sub-paths into the boot list after successfully booting the system:

   ```
   bootlist -m normal hdisk1 hdisk6 hdisk11 hdisk16
   ```

8. Enable the boot device for multipath software control.

   Note: You can skip this step for native MPIO.

- Type the following command for a PowerPath installed system:

  ```
  pprootdev on
  reboot
  ```

- Type the following command for a Veritas DMP installed system:

  ```
  vxdmpadm native enable vgname=rootvg
  reboot
  ```

  Note: Refer to the Veritas support page for more information.

.

CHAPTER 12

AIX with PowerStore Support

This chapter provides information about Dell EMC PowerStore array support in the IBM AIX environment.

Overview

PowerStore is a versatile platform with a performance-centric design that delivers multi-dimensional scale, always on data reduction, and support for next generation media. It brings the simplicity of public cloud to on-premises infrastructure, streamlining operations with a built-in machine learning engine and seamless automation, while offering predictive analytics to easily monitor, analyze, and troubleshoot the environment. PowerStore is also highly adaptable, providing the flexibility to host specialized workloads directly on the appliance and modernize infrastructure without disruption, all while offering investment protection through flexible payment solutions and data-in-place upgrades.

Prerequisites

Before configuring the PowerStore in IBM AIX environment, ensure that the following requirements are met:

Host connectivity

See the Dell EMC Simple Support Matrices or contact your Dell EMC representative for the latest information on qualified hosts, host bus adapters, and connectivity equipment.

- Supported AIX version: AIX7.1, and AIX 7.2

ODM requirement

PowerPath and Native MPIO are supported with PowerStore connect to AIX hosts, and requires minimum Dell EMC ODM version 6.2.0.0.

Note: Veritas DMP is not supported.

PowerPath requirement

The minimum supported version is PowerPath for AIX 7.0.

FC connection configuration

Installing the PowerStore ODM filesets is necessary to configure the AIX host to use PowerStore disks over Fibre Channel (FC). Different filesets are required for native MPIO and third party multipath software. Following steps provide instructions for PowerStore FC configuration.

Install the ODM kit

1. Remove all of the PowerStore disks, if they are recognized as an **Other FC SCSI Disk Drive** by the system:

   ```
   lsdev -Cc disk | grep "Other FC SCSI Disk Drive" | awk {'print $1'}
   | xargs -n1 rmdev -dl
   ```

2. Download the Dell EMC supplied AIX ODM package version 6.2.0.0 from the Dell EMC ftp server (**ftp.emc.com**).

   a. Run the following command to go to the ftp server:

      ```
      ftp ftp.emc.com
      ```

   b. Log in with an anonymous username and your email address as password.

   c. Run the following command to go to the package location:

      ```
      cd /pub/elab/aix/ODM_DEFINITIONS
      ```

   d. Run the following command to download the package to the local host:

      ```
      get DELLEMC.AIX.6.2.0.0.tar.Z
      ```

   e. Run the **quit** command to exit.

3. Place Dell EMC supplied AIX ODM package version 6.2.0.0 in the `/usr/sys/inst.images` directory for installation.

4. Run the following commands to uncompress and untar the package:

   ```
   uncompress DELL EMC.AIX. 6.2.0.0.tar.Z
   tar -xvf DELL EMC.AIX. 6.2.0.0.tar
   rm .toc
   inutoc
   ```

5. Install the desired Dell EMC PowerStore ODM filesets:

   - For PowerPath, run the following commands:

     ```
     installp -ac -gX -d . EMC.PowerStore.aix.rte
     installp -ac -gX -d . EMC.PowerStore.fcp.rte
     ```

   - For MPIO, run the following commands:

     ```
     installp -ac -gX -d . EMC.PowerStore.aix.rte
     installp -ac -gX -d . EMC.PowerStore.fcp.MPIO.rte
     ```

     Note: You can also use **smitty** for the filesets installation.

Scan and verify PowerStore LUNs

1. Run the **cfgmgr** AIX system command to scan disks.
2. Run the **lsdev -Cc disk** command to verify all FC disks are configured properly, and show the recognized PowerStore disks.
   - If using PowerPath, output similar to the following is displayed:

     ```
     hdisk1    Available    DELL EMC PowerStore FCP VRAID Disk
     hdisk2    Available    DELL EMC PowerStore FCP VRAID Disk
     ```

   - If using MPIO, output similar to the following is displayed:

     ```
     hdisk1    Available    DELL EMC PowerStore FCP MPIO VRAID Disk
     hdisk2    Available    DELL EMC PowerStore FCP MPIO VRAID Disk
     ```

Boot the host from a PowerStore disk

This section provides the requirements and the procedures for booting the host from a PowerStore disk.

Requirements

The following information provides general guidelines, requirements, and instructions for booting AIX from the PowerStore LUN. See the Dell EMC Support Matrix for supported hardware.

- If the host get stuck at IBM reference code 0554 and doesn't boot, try to configure only one logical path from PowerStore LUN to host.
- Boot from SAN can be supported for PowerStore array when FC connection is configured; iSCSI connection is not supported currently.

Set up booting from a SAN

1. Check the sub-paths of the disk, if you are using PowerPath. Record all the sub-paths.
   - Run the following command for a PowerPath installed system:

     `powermt display dev=hdiskpower#`
2. Release the sub-paths from the control of PowerPath.
   - Run the following command for a PowerPath installed system:

     `powermt remove dev=hdiskpower#`

     `rmdev -dl hdiskpower#`
3. Run the following command to clone the system into an available sub-path of the multipath disk:

   `alt_disk_install -C hdisk#`

   Note: It is recommended to configure only one sub path.

4. Run the **`lspv | grep rootvg`** command to check if an alternate installation disk was cloned successfully.

   Note: There should be an **`altinst_rootvg`** volume group.

   The following is the result of running the command:

   ```
   hdisk0          00c548453a4d0f10        rootvg                  active
   hdisk1          00c548458c4a1ab5        altinst_rootvg
   ```

5. Run the **`bootlist -om normal`** command to check the boot list.

   This shows the list that the system that has been cloned to:

   `hdisk1`
6. Reboot the system.
7. Run the following command to add sub-paths into the boot list after successfully booting the system:

```
bootlist -m normal hdisk1 hdisk6 hdisk11 hdisk16
```

8. Enable the boot device for multipath software control.

   Note: You can skip this step for native MPIO.

   - Run the following command for a PowerPath installed system:

     ```
     pprootdev on
     reboot
     ```

CHAPTER 13

Dell EMC SC Series

This chapter describes Dell EMC SC Series PowerPath configuration in the IBM AIX environment and contains support information.

Prerequisites

Perform the following prerequisites before configuring Dell EMC SC Series disk in the IBM AIX environment:

Host connectivity

Refer to the Dell EMC Simple Support Matrices or contact your Dell EMC representative for the latest information on qualified hosts, host bus adapters, and connectivity equipment.

- Supported AIX version: AIX 6.1, AIX7.1, and AIX 7.2.
- PowerPath and AIX MPIO are the only multipath solutions supported with Dell EMC SC series and AIX hosts.
- Support PowerPath version: PowerPath for AIX 6.2 or later.
- For AIX MPIO support, download the Dell Storage Software Suite for AIX from Dell.com/support.

ODM version requirement

This paper discusses AIX version 6.1 through 7.2 with Dell EMC Storage Center OS (SCOS) version 7.1.x through 7.2.x with PowerPath installed. Dell EMC SC Series FCP Disk can be supported on AIX system:

- AIX 6.1 requires EMC ODM 5.3.1.2 or later.
- AIX 7.1 and AIX 7.2 require EMC ODM 6.0.0.7 or later.

PowerPath on Dell EMC SC Series

This section contains information for PowerPath on Dell EMC SC Series.

Dell EMC currently supports PowerPath for AIX 6.2 or later for FC multipath functionality with AIX. Refer to the Dell EMC Simple Support Matrices for the latest support information.

Installing the Dell EMC ODM kit

1. Remove all the Dell EMC SC Series disks, if they're recognized as "**Other FC SCSI Disk Drive** by system.

   ```
   lsdev -Cc disk | grep "Other FC SCSI Disk Drive" | awk {'print $1'}
   | xargs -n1 rmdev -dl
   ```

2. Download the EMC-supplied AIX ODM Package at minimum version 5.3.1.2 or 6.0.0.7 from the EMC ftp server ftp.emc.com.

   f. Type the following command:

   ```
   ftp ftp.emc.com
   ```

   g. Login with a user name of anonymous and your email address as a password.

   h. Type the following command:

   ```
   cd /pub/elab/aix/ODM_DEFINITIONS
   ```

   i. Type the file transfer mode to binary:

   ```
   binary
   ```

   j. Type the file transfer mode to binary:

   ```
   get EMC.AIX.5.3.x.x.tar.Z
   ```

   k. Type the file transfer mode to binary:

   ```
   bye
   ```

3. Place the EMC Supplied AIX ODM Package into the /usr/sys/inst.images directory for installation. Uncompress and untar the package. Type the following command:

   ```
   cd /usr/sys/inst.images
   uncompress EMC.AIX.5.3.x.x.tar.Z
   tar -xvf EMC.AIX.5.3.x.x.tar
   ```

4. Install the desired Dell EMC SC Series ODM files and PowerPath 6.2 for AIX.

   ```
   rm .toc
   inutoc .
   ```

   l. Install the following two files:

   ```
   installp -agXd . EMC.DellSC.aix.rte
   installp -agXd . EMC.DellSC.fcp.rte
   installp -agXd . EMCpower
   ```

Note: You can also install the files using smitty.

Scan and verify Unity LUNs

Use the following commands:

5. Type the following AIX system command to scan diskc:

   ```
   cfgmgr
   ```

6. Type the following command:

   ```
   powermt config.
   ```

7. Verify that all disks have been configured properly. Show the recognized Dell EMC SC Series FC disks by entering:

   ```
   powermt display dev=all class=sc

   Pseudo name=hdiskpower3
   SC ID=5000D31000DD4E00 [SC4020-FC-1]
   Logical device ID=6000D31000DD4E00000000000000048C [ax202041 2]
   state=alive; policy=ADaptive; queued-IOs=0
   ==============================================================================
   ---------------- Host ---------------   - Stor -  -- I/O Path --   -- Stats ---
   ###  HW Path             I/O Paths    Interf.  Mode    State   Q-IOs  Errors
   ==============================================================================
     1 fscsi1             hdisk30   56654-07  active   alive     0     0
     1 fscsi1             hdisk20   56654-05  active   alive     0     0
     0 fscsi0             hdisk14   56654-07  active   alive     0     0
     0 fscsi0             hdisk24   56654-05  active   alive     0     0
   ```

-
- Type the **cfgmgr** AIX system command to scan disks.
- Type the **lsdev -Cc disk** command to verify that all FC disks are configured properly and show the recognized Unity disks.
- If using PowerPath/DMP, the result will be as follows:

  ```
  hdisk1      Available          EMC CLARiiON FCP VRAID Disk

  hdisk2      Available          EMC CLARiiON FCP VRAID Disk
  ```

- If using MPIO, the result will be as follows:

  ```
  hdisk1      Available          EMC CLARiiON FCP MPIO VRAID Disk

  hdisk2      Available          EMC CLARiiON FCP MPIO VRAID Disk
  ```

.

PART 3

High Availability

Part 4 includes:

CHAPTER 12

High Availability With VMAX, Symmetrix, VNX Series, or CLARiiON and AIX

This chapter discusses VMAX, Symmetrix, VNX, or CLARiiON systems in the AIX server high availability and concurrent access cluster environment.

Note: Any general references to Symmetrix or Symmetrix array include the VMAX3 Family, VMAX, and DMX.

Overview

This chapter applies to all currently supported High-Availability Cluster Multi-Processing (HACMP) installations. For more detailed and updated information, refer to the `README` file.

This chapter provides information on using AIX systems in high-availability clusters.

These systems use HACMP software to enable shared disk access and automate failover of system resources.

About HACMP

IBM HACMP software increases the high availability of storage system configurations with two servers; that is, split bus, dual-initiator, and dual-initiator/dual-bus configurations. HACMP provides server control programs that perform server failover.

Dell EMC HACMP support software includes scripts that support Fibre Channel disk arrays under the HACMP software.

You must install the software in this order:

1. Adapter drivers
2. Agent and CLI software
3. HACMP software
4. HACMP support software

The procedure in this chapter assumes that you have installed the software in that order and that you have bound LUNs.

High availability configurations

Symmetrix, VNX series, CLARiiON, and AIX systems offer a number of clustering alternatives, for both high availability (HA) and concurrent access. The following paragraphs present an outline of HA and concurrent access functions and enablers.

HACMP for AIX

Features of HACMP include:

- Up to eight nodes
- Detection of node, network adapter, and network failures (and others via AIX error notification)
- IP address takeover (network failover)
- Application failover (user-definable)
- Disk, volume group, and file system failover
- Concurrent Access (OPS support)
- JFS support
- MPIO device support
- VIO support

HACMP/ES

Features of HACMP/ES (Enhanced Scalability) include:

- Support for RS/6000, eServer pSeries, System p
- Up to 32 nodes (HACMP-dependent)
- HACWS option to fail over CWS to a standby workstation
- Detection of node, network adapter, and network failures (and others via AIX error notification)
- IP address takeover
- VIO support
- Application failover (user definable)
- Disk, volume group, and file system failover
- MPIO device support
- Fast Disk Takeover
- Heartbeat over disk
- Concurrent access (OPS support)
- Support for JFS/JFS2

VMAX, Symmetrix, VNX series, and CLARiiON in the HACMP cluster

IBM HACMP is cluster software for RS/6000, eServer pSeries, and System p. Symmetrix systems are supported with these combinations of HACMP and AIX:

- PowerHA 5.5 with AIX 5.3
- PowerHA 5.5 with AIX 6.1
- PowerHA 5.5 with AIX 7.1
- PowerHA 6.1 with AIX 5.3
- PowerHA 6.1 with AIX 6.1
- PowerHA 6.1 with AIX 7.1
- PowerHA 7.1 with AIX 6.1
- PowerHA 7.1 with AIX 7.1

The use of Cluster Aware AIX (CAA) for integrated topology management is new with PowerHA 7.1. A major part of Cluster Aware AIX (CAA) is the central repository. The central repository is stored on a dedicated disk, shared between all participating nodes.

Note: Requires minimum 7100-01-00-1140 or 6100-07-00-1140 to use PowerPath as the central repository disk.

Before managing a HACMP system, be sure to review the Dell EMC Simple Support Matrices and read the release notes that accompany the software for required APARs and notes specific to HACMP versions.

EMC highly recommends that you complete the HACMP training course before attempting any configuration procedures. For more information about specific courses, contact IBM Educational Services. For detailed guidelines on HACMP cluster administration, refer to the HACMP documentation set.

HACMP high availability

High availability of a properly planned and maintained cluster is ensured by HACMP software running on each node. HACMP *daemons* detect failures of a node, network, or network adapter and respond with appropriate recovery or failover actions. Applications and other resources can be protected by a failover process that moves an application or service from a failed node to a surviving node. Applications are installed on each node and disks are shared, so any node can run the application and access its data.

HACMP can be used with AIX LVM to provide highly available disks, mirrored disks, and fast file recovery using a journaled file system (JFS).

Resource groups

Resource groups defined by HACMP can include applications, file systems, volume groups, raw disk PVID, and IP address resources. Resource groups are also defined by their node relationships as part of the configuration process. Node relationship can be defined as one of three types:

- **Cascading** — Resources fail over to surviving nodes as determined by defined priority. Nodes rejoining the cluster regain control of resources for which it has the highest priority.
- **Rotating** — Resources rotate among nodes. Failover is to the node with the highest defined priority for the resource. However, a node rejoining the cluster does not reacquire its resources, but instead acts as a standby.
- **Concurrent access** — Concurrent access resources are shared by multiple nodes. All nodes own the resource, and have access to it under control of the lock manager. No priorities are assigned.

Note: Only raw logical volumes may be defined as a concurrent access resource. File systems and JFS are not supported.

Starting with HACMP 5.2, the policies in Table 13 exist for resource group startup, fallover, and fallback.

Table 13 Resource group startup, fallover, and fallback policies

Startup	• **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. • **Online on All Available Notes.** The resource group is brought online on all nodes. • **Online Using Distribution Policy.** Only one resource group is brought online on a node, or on a node per network, depending on the distribution policy specified (node or network).
Fallover	• **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 node list. • **Fallover Using Dynamic Node Priority.** Before selecting this option, configure a dynamic node priority policy that you want to use. Or you can select one of the three predefined dynamic node priority policies. • **Bring Offline (on Error Node Only).** Select this option to bring a resource group offline on a node during an error condition.
Fallback	• **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. If you do not configure a delayed fallback policy, the resource group falls back immediately when a higher priority node joins the cluster. • **Never Fallback.** A resource group does *not* fall back when a higher priority node joins the cluster.

Configuration types

The HACMP cluster configuration is determined by the number of nodes and the way the resources are defined. In the following examples, the configuration possibilities are described using the simplest two-node cluster; however, up to eight nodes are supported with the HAS and CRM features, while up to 32 nodes are supported with the ES and ESCRM features:

- **Hot standby** — In this configuration, all resources are cascading, with a single node having the highest priority among them. One node is normally idle, waiting to take over if the *owning* node (the node that *owns* the resources) fails.

  If the owning node fails, the standby node takes over the resources. When the failed node rejoins the cluster, the resources are returned to it and it again becomes the owning node.

- **Rotating standby** — This configuration is identical to hot standby, except that when the failed node rejoins the cluster, the resources are not returned to it unless the standby node fails.

- **Mutual takeover** — In this configuration, resources are cascading and divided among the nodes. Certain resources are defined as *owned* by each node. If either node fails, the other node takes over all resources. When the failed node rejoins the cluster, the resources are returned to it and it again becomes the owning node.

- **Concurrent access** — In this configuration, two nodes are active simultaneously, sharing the same physical disk resources. (The disk resources are considered *concurrent*.) Any other resources are divided between the two nodes, each owning some of them. The

resources not owned by each node are designated as *cascading*. If either node fails, the other node takes over all resources. When the failed node rejoins the cluster, the resources originally assigned to it are returned to it.

AIX 5L v5.1 and up provides a new form of concurrent mode: enhanced concurrent mode. Concurrent volume groups are created as enhanced concurrent mode volume groups by default. For enhanced concurrent volume groups, the Concurrent Logical Volume Manager (CLVM) coordinates changes between nodes through the Group Services component of the Reliable Scalable Cluster Technology (RSCT) facility in AIX. Group Services protocols flow over the communications links among the cluster nodes.

Interhost networks

All nodes of the HACMP cluster are connected to at least one common network (Ethernet, token-ring, FDDI, or ATM). The network interface is used to monitor heartbeat messages and pass cluster control messages between nodes.

Multiple networks and network adapters may be used in a primary and backup arrangement. HACMP can fail over from a primary adapter to a backup adapter on the same network and host.

An RS-232 serial network or disk heartbeat (over an enhanced concurrent mode disk) also can be configured to provide additional backup for interhost communication.

Configuring heartbeating over disk

You can configure disk heartbeat 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. Heartbeat networks contain:

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

The following considerations apply:

- To check the disk load, use the **`filemon`** command. If the disk is heavily used, it may be necessary to change the tuning parameters for the network to allow more missed heartbeats. Disk heartbeat networks use the same tuning parameters as RS232 networks.
- To check the disk load, use the **`filemon`** command. If the disk is heavily used, it may be necessary to change the tuning parameters for the network to allow more missed heartbeats. Disk heartbeat networks use the same tuning parameters as RS232 networks.
- To configure a heartbeat over disk network:

  1. Ensure that the disk is in an enhanced concurrent mode volume group (this can be done using CSPOC).
  2. Create a diskhb network.
  3. Add disk-node pairs to the diskhb network.

```
   /usr/es/sbin/cluster/utilities/claddnode -a'hdisknobdeb'
:'diskhb' :'net_diskhb_01' :serial:service:'/dev/hdiskpower4'
-n'nodea'
```

```
   /usr/es/sbin/cluster/utilities/claddnode -a'hdisknobdea'
:'diskhb' :'net_diskhb_01' :serial:service:'/dev/hdiskpower4'
-n'nodeb'
```

Ensure that each node is paired with the same disk as identified by the device name, for example, hdiskpower1 and PVID.

4. Synchronize the cluster.

Verifying disk heartbeat configuration

Cluster verification verifies whether a disk heartbeat network is configured correctly. RSCT logs provide information for disk heartbeat networks that is similar for information for other types of networks.

When you run verification, the **clverify** utility verifies that:

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

The first step in troubleshooting a disk heartbeat network is to test the connections. The **dhb_read** command tests connectivity for a disk heartbeat network.

To use **dhb_read** to test a disk heartbeat connection:

1. Set one node to run the command in data mode:

   /usr/sbin/rsct/bin/dhb_read -p hdisk# -r

   where `hdisk#` identifies the hdisk in the network, such as hdiskpower4, hdisk1, or hdiskpower1.

   For example:

   ```
   /usr/sbin/rsct/bin/dhb_read -p hdiskpower4 -r
   ```

2. Set the other node to run the command in transmit mode:

   /usr/sbin/rsct/bin/dhb_read -p hdisk# -t

   where `hdisk#` identifies the hdisk in the network. The `hdisk#` is the same on both nodes.

   For example:

   ```
   /usr/sbin/rsct/bin/dhb_read -p hdiskpower4 -t
   ```

3. The following message indicates that the communication path is operational:

   ```
   Link operating normally
   ```

   For example:

```
(255)nodea/> /usr/sbin/rsct/bin/dhb_read -p hdiskpower4 -r
Receive Mode:
Waiting for response ...
```

```
Link operating normally

(130)nodeb/> /usr/sbin/rsct/bin/dhb_read -p hdiskpower4 -t
Transmit Mode:
Waiting for response ...
Link operating normally
```

Adding support for MPIO devices

Support for Symmetrix-based MPIO devices requires ODM Kit fileset `devices.pci.mpio 5.2.0.1` or higher. This specific ODM is required for use with MPIO and Symmetrix (whether HACMP is being utilized or not). When a shared volume group in an HACMP cluster is accessed through MPIO, it must be defined as an enhanced concurrent volume group. It is important to note that this does not limit the use of MPIO with HACMP to concurrent resource groups. Starting with HACMP 5.1, and continuing with HACMP 5.2, enhanced concurrent mode volume groups can be used with non-concurrent resource groups. This facility, referred to as "Fast Disk Takeover" in the HACMP publications, combines the recovery speed of concurrent volume groups with the secure, single system access of non-concurrent volume groups. However, when this facility is used, it is recommended that the customer employ zoning or physical connectivity to ensure that only systems that are part of the HACMP cluster have access to the volume group.

In the case where these configuration guidelines are adhered to, it is permissible for the Symmetrix LUNs controlled by HACMP for the customer to change using the **chdev** command, the reservation setting from `single_path` to `no_reserve`, and the algorithm to `round_robin`. These changes will allow MPIO to enable load balancing (in addition to failover) for the LUNs in question, potentially yielding enhanced performance.

1. Determine which MPIO devices will get the load balancing attributes set.

```
# lsattr -El hdisk2
PCM               PCM/friend/MSYMM_RAID1        Path Control Module                 True
PR_key_value      none                          Persistant Reserve Key Value        True
algorithm         fail_over                     Algorithm                           True
clr_q             yes                           Device CLEARS its Queue on error    True
hcheck_interval   10                            Health Check Interval               True
hcheck_mode       nonactive                     Health Check Mode                   True
location                                        Location Label                  True
lun_id            0x57c000000000000             Logical Unit Number ID          False
lun_reset_spt     yes                           FC Forced Open LUN              True
max_transfer      0x40000                       Maximum TRANSFER Size           True
node_name         0x5006048accc82def            FC Node Name                    False
pvid          00002dffbfa8d5d20000000000000000  Physical volume identifier      False
q_err             no                            Use QERR bit                    True
q_type            simple                        Queue TYPE                      True
queue_depth       16                            Queue DEPTH                     True
reserve_policy    single_path                   Reserve Policy                  True
rw_timeout        40                            READ/WRITE time out value       True
scsi_id           0x6a1a13                      SCSI ID                         False
start_timeout     180                           START UNIT time out value       True
ww_name           0x5006048accc82def            FC World Wide Name              False
```

2. Modify the settings for reserve_policy.

```
# chdev -l hdisk2 -a reserve_policy=no_reserve
hdisk2 changed
```

3. Modify the settings for algorithm.

```
# chdev -l hdisk2 -a algorithm=round_robin
hdisk2 changed
```

4. Verify your changes.

```
# lsattr -El hdisk2
```

```
PCM               PCM/friend/MSYMM_RAID1            Path Control Module              True
PR_key_value      none                              Persistant Reserve Key Value     True
algorithm         round_robin                       Algorithm                        True
clr_q             yes                               Device CLEARS its Queue on error True
hcheck_interval   10                                Health Check Interval            True
hcheck_mode       nonactive                         Health Check Mode                True
location                                            Location Label                   True
lun_id            0x57c000000000000                 Logical Unit Number ID           False
lun_reset_spt     yes                               FC Forced Open LUN               True
max_transfer      0x40000                           Maximum TRANSFER Size            True
node_name         0x5006048accc82def                FC Node Name                     False
pvid              00002dffbfa8d5d20000000000000000  Physical volume identifier       False
q_err             no                                Use QERR bit                     True
q_type            simple                            Queue TYPE                       True
queue_depth       16                                Queue DEPTH                      True
reserve_policy    no_reserve                        Reserve Policy                   True
rw_timeout        40                                READ/WRITE time out value        True
scsi_id           0x6a1a13                          SCSI ID                          False
start_timeout     180                               START UNIT time out value        True
ww_name           0x5006048accc82def                FC World Wide Name               False
```

Note: Review the Dell EMC Support Matrices for operating system and HACMP software requirements. For MPIO configurations using load balancing, the reservation setting `no_reserve` and the algorithm `round_robin` is supported on **`rootvg`**.

HACMP administration

HACMP is administered through command line or through the `SMIT` HACMP menus. This section provides an overview of utilities provided for maintenance and monitoring of the HACMP cluster.

Information provided here is intended to help familiarize you with tasks and commands associated with managing Symmetrix storage in the HACMP cluster. Refer to the *HACMP Administration Guide* for detailed descriptions and guidelines for planning and implementing HACMP administration.

System Resource Controller

The System Resource Controller (SRC) is an AIX facility that controls the HACMP daemons. The following subsystems are defined in the cluster group under SRC:

- **Cluster Manager** (`clstrmgr`) — Monitors status, reacts to events, and maintains heartbeat.
- **Cluster Information Program** (`clinfo`) — Optional daemon provides status to clients.
- **Cluster SMUX Peer** (`clsmuxpd`) — Maintains status of cluster objects. Works with the SNMP daemon (`snmpd`).
- **Cluster Lock Manager** (`cllockd`) — Required for concurrent access nodes only in the SRC lock group.

Note: The SRC is accessed through `SMIT` screens to start, stop, or monitor the cluster services.

Note: Cluster Lock Manager on HACMP 5.1 is not supported on 64-bit kernels and is no longer supported in HACMP 5.2 or higher.

Starting cluster services

Cluster nodes are started one at a time after all have been booted. If the cluster is configured for cascading or rotating resources, be sure to start nodes in the correct sequence.

To access the `SMIT` **Start Cluster Services** screen, log in as `root` and enter:

```
smit clstart
```

The startup screen contains the following entry fields:

Start now, on system restart or both

BROADCAST message at startup?

Startup Cluster Lock Services?

Startup Cluster Information Services?

Enter your selection for each field, and press **Enter** to activate the cluster services defined on this node. Monitor the messages displayed when the `node_up` script is started and when the `node_up_complete` script is finished. Any error messages will be logged in `/var/adm/cluster.log` or `/tmp/hacmp.out`. The startup process is repeated on all nodes.

Stopping cluster services

Cluster services are stopped on one or more nodes to perform scheduled maintenance or to make changes to the configuration. Stopping a cluster node should be planned to minimize impact on high availability services. If all nodes are to be stopped (required for changes to the ODM cluster information), stop one node at a time.

Do not use the **kill -9** command to stop the cluster manager. This causes an abnormal exit of **clstrmgr**, resulting in a system halt and failover to remaining nodes. Use the **graceful with takeover** stop if you want to intentionally initiate failover.

The following types of cluster stops are available:

- **Graceful** — Applications are shut down and resources released. Resources are not taken over by other nodes.
- **Graceful with Takeover** — Applications are shut down and resources released. Resources are taken over by surviving nodes.
- **Forced** — HACMP software is stopped immediately and no HACMP scripts are run. Other nodes view this as a graceful shutdown and do not attempt takeover. Resources are maintained and applications that do not require the HACMP daemons can continue uninterrupted.

To access the `SMIT` **Stop Cluster Services** screen, log in as `root` and enter:

```
smit clstop
```

The stop screen contains the following entry fields:

Start now, on system restart or both

BROADCAST message at startup?

Shutdown mode (graceful, graceful with takeover, forced)

Enter your selections for each field, and press **Enter** to stop the cluster services defined on this node. This process may be repeated on other nodes to stop the entire cluster.

Monitoring cluster services

HACMP provides a utility, `/usr/bin/cluster/clstat`, to report status of various cluster components. The `clstat` utility is a Cluster Information client; `clinfo` must be running for this utility to function correctly. Cluster status information is presented either graphically in an X-Windows display or as an ASCII display. Refer to the *HACMP Administration Guide* and `clstat` main pages for more information.

The status of the HACMP daemons can be monitored using the `SMIT` **Show Cluster Service**s screen. To access the screen, log in as `root` and enter:

```
smit clshow
```

Identifying ghost disks

Ghost disks are duplicate profiles of the disks created during disk configuration processing. Ghost disks must be removed to allow AIX to vary on the volume group residing on the original disks.

A ghost disk can be created when the disk configuration method is run while a disk is being reserved by another node. In this case, the configuration method cannot read the PVID of the reserved disk to identify it as an existing disk.

This can happen when:

- A node fails
- A node fails.
- HACMP takes over the failed disks. During the takeover, the other node breaks any outstanding disk reserve on the failed node and establishes its own reserve on the disks.
- The failed node reboots. The disk configuration method run at boot time cannot read the disks due to the disk reserve held by another node.

Ghost disks are a problem for two reasons:

- The presence of superfluous disk profiles is confusing to the system.
- If the ghost disk is in an available state and the PVID is returned by `lsdev`, then any attempt to make the real disk available will fail, since the two disks refer to the same device. This prevents use of the real disk for operations like `varyon`.

Identifying ghost disk is supported in HACMP in the following way, and is accessed via `smit` under custom disk methods:

1. The method is passed as an input hdisk name, such as `hdisk42`.
2. The method writes to standard output a list of disk names that are ghost disks for the given disk.
3. If there are no ghost disks for the given disk, nothing is returned.

A return code of `0` indicates success; any other value will cause processing to be terminated for the given disk.

HACMP supports values of SCSI2 and SCSI3 for the name of this method. If either of those values are specified, HACMP uses the existing disk processing that is being used to identify ghost disks for IBM SCSI-2 or SCSI-3 disks:

- SCSI2 ghost disks are those with different names from the given disk, but identical parent and location attributes in `CuDv`.
- SCSI3 ghost disks are those with different names from the given disk, but identical parent and location attributes in `CuDv`; and identical **`lun_id, scsi_id`**, and **`ww_name`** attributes in `CuAt`.

If you are using PowerPath version 3.0.3 or higher and HACMP 5.3 or below, it is required that you implement the custom “emcpowerreset” binary for clearing disk reservations.

Implementing emcpowerreset

Before implementing **emcpowerreset** the following must apply:

- In a PSSP/VSD environment (version 3.4 and 3.5) with AIX 5.1 and higher and PowerPath 3.0.3 and higher do not use the **emcpowerreset** utility.
- If upgrading an existing HACMP environment from AIX 4.3.x, PP 3.0.2 and below, or a non-PowerPath environment, modify the **disktype.lst** file to remove any lines added to enable LUN reset. See EMC Knowledgebase solution emc27897 for more information.
- If PP 4.2 and higher is installed a new (Rev 2) version of the **emcpowerreset** binary is available. The new version is backward compatible with PP 3.x.
- SCSI attach is no longer supported with PP 4.2 and higher.
- All Symmetrix arrays are assumed to be at Enginuity levels 5267.26.18s / 5567.33.18s or 5268.05.06S / 5568.05.05S and higher. These levels are required to support LUN reset.
- In the examples below, it is assumed that the **emcpowerreset** binary has been copied to the directory `/usr/lpp/EMC/Symmetrix/bin` for Symmetrix attach and `/usr/lpp/EMC/CLARiiON /bin` for VNX series or CLARiiON attach. If this is not the case, simply change the path to reflect the correct location.
- If implementing HACMP on VNX series or CLARiiON LUNs with PowerPath either the set_scsi_id script or the cfgscsi_id executable is required. Refer to EMC Knowledgebase solution EMC97234 for details on the set_scsi_id script and EMC Knowledgebase solution emc143075 for details on the cfgscsi_id executable. This does not apply to HACMP 5.4 when using the IBM HACMP native cl_fscsilunreset utility.
- Because **emcpowerreset** can take approximately 20 seconds per device to break reservations, Dell EMC recommends setting the break reserves in parallel to true to improve fail-over times. See Dell EMC Knowledgebase solution emc122557 for additional details.

IMPORTANT

Starting with HACMP 5.4, the IBM HACMP native cl_fscsilunreset utility correctly clears disk reservations for Dell EMC devices and the installation of **emcpowerreset**, **set_scsi_id**, and `cfgscsi_id` utilities are no longer required, but are still highly recommended by both Dell EMC and IBM.

Dell EMC highly recommends using emcpowerreset, set_scsi_id, and cfgscsi_id with HACMP 5.4 and later to reduce longer cluster startup and failover times (refer to EMC Knowledgebase Solution emc156011 and EMC Knowledgebase Solution emc69100) and to avoid running into a rare race condition that could affect cluster failovers that require disk reservations to be broken.

PowerHA SystemMirror

This section contains the following information:

Cluster types

Two new cluster types are introduced in PowerHA 7.1.2 Enterprise Edition: stretched and linked.

- In a stretched cluster:
  - Sites defined are in the same geographical location.
  - Clusters share repository disk between sites.
  - Multicast is required within the cluster.

  Figure 12 shows an example of a stretched cluster.

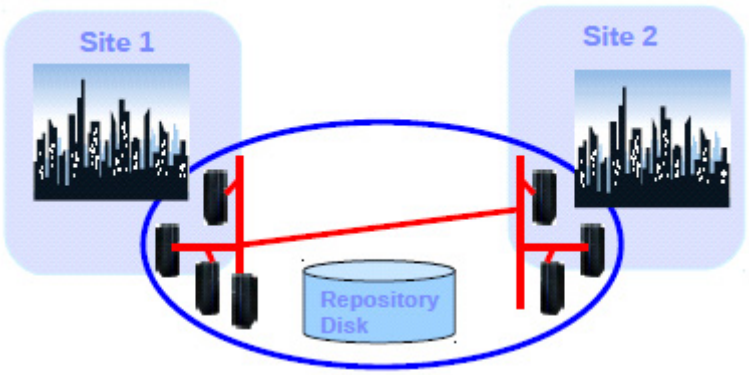


Figure 12 Stretched cluster example

- In a linked cluster:
  - Sites are located at different geographical locations.
  - Each cluster has its stand-alone repository disk.
  - Unicast is used to communicate between two sites

  Figure 13 shows an example of a stretched cluster.

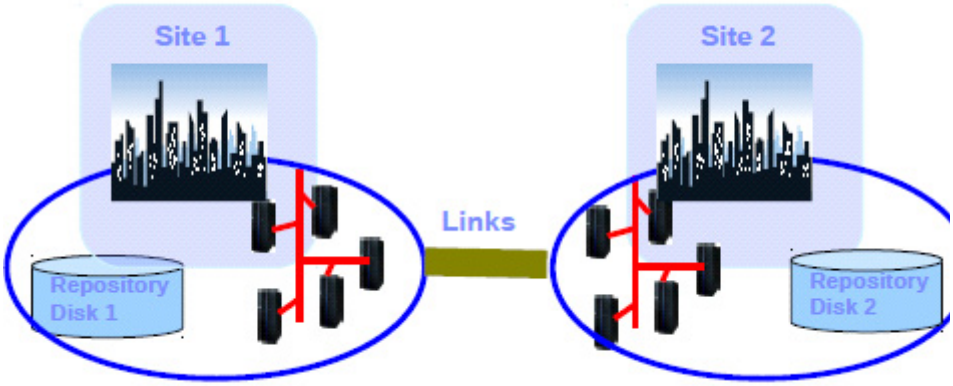


Figure 13 Linked cluster example

PowerHA systemMirror 7.1.2 uses the Cluster Aware AIX (CAA) services to configure, verify and monitor the cluster topology.

Stretched clusters use multicast communication between all nodes across all sites. In this case, a single multicast IP address is used in the cluster. CAA is unaware of the PowerHA site definition.

Linked clusters use multicast communication only within the site and unicast protocol for communication between the cluster nodes across the sites. Unlike multicast, unicast communication involves a one-to-one communication path with one sender and one receiver. This cluster type uses two multicast IP group addresses, one in each site. For the particular case of a linked cluster with two nodes, one at each site, the multicast addresses are defined, but the node-to-node communication is using only unicast.

For PowerHA operation, the network infrastructure must handle the IP multicast traffic properly:

- Enable multicast traffic on all switches used by the cluster nodes.
- Check the available multicast traffic IP address allocation.
- Ensure that the multicast traffic is properly forwarded by the network infrastructure (firewalls, routers) between the cluster nodes, according to your cluster type requirements (stretched or linked).

PowerHA basic cluster configuration

This section will describe how to set up a basic PowerHA cluster:

Configuring the network

1. Set the IP address of two Virtual Ethernet Adapters on each node by using **smitty mktcpip**.

   For example, the following is the network setup for one node aix119057.

   ```
   aix119057/> smit mktcpip
   ```

   Configure en0 and en1 separately. There is no gateway setting at this time.

```
en0:
* HOSTNAME                                           [aix119057]
* Internet ADDRESS (dotted decimal)                  [192.168.1.57]
  Network MASK (dotted decimal)                      [255.255.255.0]
* Network INTERFACE                                   en0
  NAMESERVER
           Internet ADDRESS (dotted decimal)         []
           DOMAIN Name                               []
  Default Gateway
       Address (dotted decimal or symbolic name)     []
       Cost                                          [0]                                #
       Do Active Dead Gateway Detection?             no                                +
  Your CABLE Type                                    N/A                               +
  START TCP/IP daemons Now                           no                                +

en1:
* HOSTNAME                                           [aix119057]
* Internet ADDRESS (dotted decimal)                  [192.168.2.57]
  Network MASK (dotted decimal)                      [255.255.255.0]
* Network INTERFACE                                   en1
  NAMESERVER
           Internet ADDRESS (dotted decimal)         []
           DOMAIN Name                               []
  Default Gateway
       Address (dotted decimal or symbolic name)     []
       Cost                                          [0]                          #
       Do Active Dead Gateway Detection?             no                                +
  Your CABLE Type                                    N/A                               +
  START TCP/IP daemons Now                           no                                +
```

2. Configure the persistent IP using `smit`.

```
aix119057/> smit inetalias
Add an IPV4 Network Alias then fill the filed.
  Network INTERFACE                                   en0
* IPV4 ADDRESS (dotted decimal)                      [10.108.119.57]
  Network MASK (hexadecimal or dotted decimal)       []
```

3. Set the default gateway entry in any of the base interfaces using `smit`.

```
aix119057/> smit mktcpip
```

```
Default Gateway
       Address (dotted decimal or symbolic name)      [10.108.119.1]
```

Configuring name resolutions

Before creating the cluster, you must perform the following tasks.

1. Edit `/etc/environment` and add `/usr/es/sbin/cluster/utilities` and `/usr/es/sbin/cluster/` to the `$PATH` variable.
2. Populate `/etc/cluster/rhosts`. For example, example for node `aix119057`:

```
aix119057/> cat /etc/hosts
192.168.1.57    aix119057b1
192.168.1.58    aix119058b1
192.168.2.58    aix119058b2
192.168.1.53    aix119053b1
192.168.2.53    aix119053b2
10.108.119.53   aix119053
10.108.119.57   aix119057
10.108.119.58   aix119058
192.168.2.57    aix119057b2
10.108.119.55   appsvc

aix119057/> cat /etc/cluster/rhosts
10.108.119.53
10.108.119.55
10.108.119.57
10.108.119.58
192.168.1.53
192.168.2.53
192.168.1.57
192.168.2.57
192.168.1.58
192.168.2.58
```

3. After editing the rhost file, restart the clcomd daemon using the **`stopsrc -s clcomd`** and the **`startsrc -s clcomd`** commands, respectively.
4. Run the following command to check the clcomd status.

```
aix119057/> lssrc -s clcomd
Subsystem         Group            PID          Status
 clcomd           caa              6291650      active
```

Installing the EMC ODM support package for AIX

Install the IBM Fibre Channel Interface Support software (ODM) on the AIX server. The ODM software enables the host operating system to recognize devices as belonging to a Symmetrix, CLARiiON, or VNX storage system.

1. Download the most recent EMC ODM Support package for the AIX version from the TP site.

```
ftp://ftp.EMC.com/pub/elab/aix/ODM_DEFINITIONS
```

2. Uncompress and untar the fileset.

```
uncompress EMC.AIX.X.X.X.tar.Z
tar -xvf EMC.AIX.X.X.X.tar
```

3. Install the Dell EMC Supplied AIX ODM Package using the AIX **`installp`** command.

```
smitty installp
```

Refer to the ODM Package README file for specific installation instructions.

Creating a cluster

Once the above steps are complete, the cluster can be created using smitty.

In the following example, a stretched cluster and a linked cluster will be created.

1. Configure the cluster through **`smitty sysmirror`** **> Cluster Nodes and Networks > Multi Site Cluster Deployment > Setup a Cluster > Nodes and Networks**.

```
Setup Cluster, Sites, Nodes and Networks
Type or select values in entry fields.
Press Enter AFTER making all desired changes.
[Entry Fields]
* Cluster Name [mycluster]
* Site 1 Name [Site1]
* New Nodes (via selected communication paths) [aix119053] +
* Site 2 Name [Site2]
* New Nodes (via selected communication paths) [aix119057 aix119058] +
Cluster Type [Linked Cluster] +
```

2. Press **Enter**.

   A discovery process is executed. This automatically creates the cluster networks and saves all the shared disks.

3. Ensure that there is a logical unit number (LUN) allocated to both AIX systems for use as the CAA repository disk.

   As an example, the following is an overview of stretched and linked cluster scenarios.

   Stretched cluster:

   Site1: `aix119057`
   Site2: `aix119058`

   Each node has five disks.

```
(0)aix119057/> lspv
hdisk0          00f6bbd54835a482                    rootvg          active
hdisk1          00f6bbd5bf8ba0d1                    None
hdisk2          00f6bbd5f34812a1                    None
hdisk3          00f6bbd54fdba0df                    None
hdisk4          00f6bbd4bf27c5ab                    None
hdisk5          00f6bbd4bf1ff5f0                    None

(0)aix119058/> lspv
hdisk0          00f6bbd5482c4b16                    rootvg          active
hdisk1         00f6bbd5bf8ba0d1                     None             active
hdisk2          00f6bbd54b7fdcea                    None
hdisk3          00f6bbd54fdba9de                    None
hdisk4          00f6bbd4bf27c5ab                    None
hdisk5          00f6bbd4bf1ff5f0                    None
```

   Notes:

   - In this case, all nodes have access to **`hdisk1`**. Therefore, it could be configured as a shared repository disk of these nodes.
   - Repository disks cannot be mirrored using AIX LVM. Therefore, it is strongly advised to have it RAID-protected by a redundant and highly available storage configuration.
   - In the event the repository disk fails or becomes inaccessible by one node or more, the node stays online but cluster configuration changes cannot be performed in this state.

- The repository disk can be replaced if it fails. The backup disk must be the same size and type as the primary disk, but can be in a different physical storage.

Linked cluster:

Site1: `aix119053`
Site2: `aixaix119057, aix119058`

Each node has five disks.

```
(0)aix119053/> lspv
hdisk0          00f6bbd447ed6f1b                    rootvg          active
hdisk1          00f6bbd4f347d521                    None
hdisk2          00f6bbd44b7f18cc                    None
hdisk3          00f6bbd44b7f1947                    None
hdisk4          00f6bbd4bf27c5ab                    None
hdisk5          00f6bbd4bf1ff5f0                    None

(0)aix119057/> lspv
hdisk0          00f6bbd54835a482                    rootvg          active
hdisk1          00f6bbd5bf8ba0d1                    None
hdisk2          00f6bbd5f34812a1                    None
hdisk3          00f6bbd54fdba0df                    None
hdisk4          00f6bbd4bf27c5ab                    None
hdisk5          00f6bbd4bf1ff5f0                    None

(0)aix119058/> lspv
hdisk0          00f6bbd5482c4b16                    rootvg          active
hdisk1         00f6bbd5bf8ba0d1                     None
hdisk2          00f6bbd54b7fdcea                    None
hdisk3          00f6bbd54fdba9de                    None
hdisk4          00f6bbd4bf27c5ab                    None
hdisk5          00f6bbd4bf1ff5f0                    None
```

Notes:

- In this case, hdisk1 from `aix119053` would be configured as repository disk of Site1. Nodes aix`119057` & `aix119058` share a common repository disk `hdisk1`.
- The repositories between sites are kept in sync internally by CAA.

In the above example, it is clear that hdisk1 is a free disk on each node. Therefore, this can be used for the repository. Next, modify the cluster definition to include the cluster repository disk. The free disk on both nodes is hdisk1.

This can be done through **`smitty sysmirror`** **> Cluster Nodes and Networks > Multi Site Cluster Deployment > Define Repository Disk and Cluster IP Address**.

For stretched cluster, only one shared repository disk needs to be defined.

```
Define Repository Disk and Cluster IP Address
Type or select values in entry fields.
Press Enter AFTER making all desired changes.
[Entry Fields]
* Cluster Name mycluster
* Repository Disk [hdisk1] +
Cluster IP Address []
```

However, for the linked cluster, each site owns its separate repository disk.

```
Multi Site with Linked Clusters Configuration
Type or select values in entry fields.
Press Enter AFTER making all desired changes.
[Entry Fields]
* Site Name Site1
* Repository Disk [(00f6bbd4f347d521)] +
```

```
Site Multicast Address [Default]
* Site Name Site2
* Repository Disk [(00f6bbd5bf8ba0d1)] +
Site Multicast Address [Default]
```

If no specific multicast is entered for the cluster IP address, PowerHA creates a valid one automatically.

4. Verify and synchronize the cluster through **`smitty sysmirror`** > **Cluster Nodes and Networks** > **Verify and Synchronize Cluster Configuration**. A basic cluster is created.

A basic cluster has now been configured. The remaining steps for a cluster configuration, such as configuring a resource group, have not changed so these will not be covered in this section.

Communication paths

This section includes the important process of sending and processing the cluster heartbeats by each participant node. Cluster communication is achieved by communicating over multiple redundant paths. The following redundant paths provide a robust clustering foundation that is less prone to cluster partitioning:

- **TCP/IP**

  PowerHA SystemMirror and Cluster Aware AIX, via multicast, use all network interfaces that are available for cluster communication. All of these interfaces are discovered by default and used for health management and other cluster communication.

- **SAN-based**

  In PowerHA 7.1, the use of heartbeat disks is no longer supported, and configuring heartbeat over SAN is the supported method to use in place of heartbeat disks. Details are further discussed in "SAN-based heartbeat configuration" on page 321.

- **Repository disk**

  Health and other cluster communication are also achieved through the central repository disk.

Normally, repository disk is in `UP RESTRICTED AIXCONTROLLED` state when TPC/IP or SAN-based heartbeat functions properly. It does not communicate with participant nodes but works as a standby path.

When TCP/IP and SAN-based paths both become unavailable, the state of repository disk switches to `UP AIX_CONTROLLED` and start to transmit heartbeat messages.

However, none of these communication paths can be used between two sites in a linked cluster. In this case, you have to use unicast communication between the sites.

SAN-based heartbeat configuration

Generally, there are two scenarios for SAN-based heartbeat configuration, each discussed further in this section:

- "Nodes using physical I/O server" on page 321
- "Nodes using Virtual I/O server" on page 322

Nodes using physical I/O server

For the first scenario, the configuration is very simple, as shown in Figure 14.

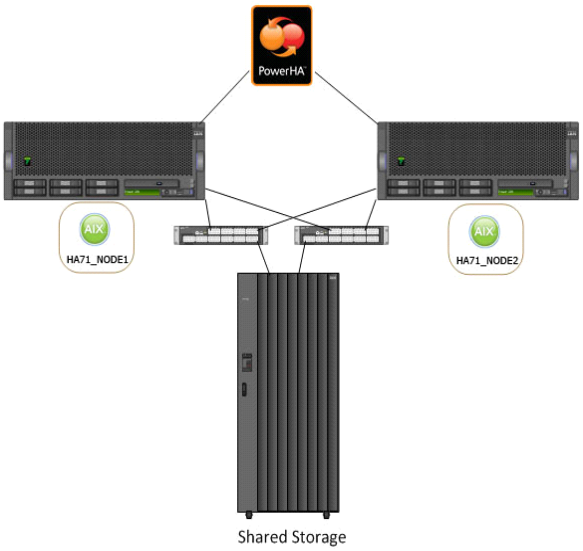


Figure 14 Nodes using physical I/O server example

Besides the storage zone, another heartbeat zone needs to be created in switch. Ensure that you zone one port from each FC adapter on the first node to another port on each FC adapter on the second node.

After performing those two zonings on the switch, the next step is to enable target mode on each of the adapter device in AIX. This needs to be performed on each adapter that has been used for a heartbeat zone.

For target mode to be enabled, both **dyntrk** (dynamic tracking) and **fast_fail** need to be enabled on the **fscsi** device, and target mode need to be enabled on the fcs device.

For example:

```
aix119057/> chdev -l fscsi0 -a dyntrk=yes -a fc_err_recov=fast_fail
fscsi0 changed
aix119057/> chdev -l fcs0 -a tme=yes
fscsi0 changed

The attribute can be checked by lsattr command after modified.

aix119057/> lsattr -El fscsi0
attach       switch    How this adapter is CONNECTED         False
dyntrk       yes       Dynamic Tracking of FC Devices        True
fc_err_recov fast_fail FC Fabric Event Error RECOVERY Policy True
scsi_id      0x6f0302  Adapter SCSI ID                       False
sw_fc_class  3         FC Class for Fabric                   True
```

Now, check whether sfwcomm devices are available. These devices are used for the PowerHA error detection and heartbeat over SAN.

```
aix119057/> lsdev -C |grep sfwcomm
sfwcomm0       Available C9-T1-01-FF Fibre Channel Storage Framework Comm
sfwcomm1       Available 10-T1-01-FF Fibre Channel Storage Framework Comm
```

Nodes using Virtual I/O server

For this scenario, storage zone and heartbeat zone also need to be configured. What differs from the physical I/O scenario is that the FC ports of the Virtual I/O Server must be zoned together in heartbeat zone, as shown in Figure 15.

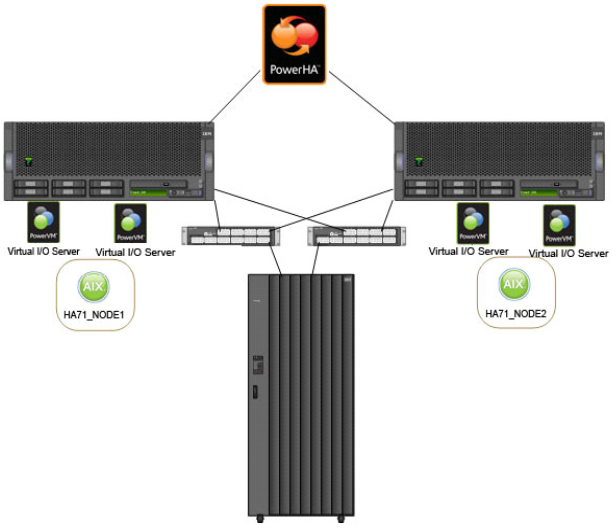


Figure 15 Nodes using Virtual I/O server example

First, check the **npiv** mapping on each VIO server.

```
VIO server 1
$ lsmap -all -npiv
Name          Physloc                            ClntID ClntName       ClntOS
------------- ---------------------------------- ------ -------------- -------
vfchost0      U8205.E6B.10BBD5P-V2-C5                 3 aix119058-HA   AIX

Status:LOGGED_IN
FC name:fcs1                    FC loc code:U78AA.001.WZSG5HN-P1-C5-T2
Ports logged in:3
Flags:a<LOGGED_IN,STRIP_MERGE>
VFC client name:fcs0            VFC client DRC:U8205.E6B.10BBD5P-V3-C5

Name          Physloc                            ClntID ClntName       ClntOS
------------- ---------------------------------- ------ -------------- -------
vfchost1      U8205.E6B.10BBD5P-V2-C6                 3 aix119058-HA   AIX

Status:LOGGED_IN
FC name:fcs3                    FC loc code:U78AA.001.WZSG5HN-P1-C7-T2
Ports logged in:3
Flags:a<LOGGED_IN,STRIP_MERGE>
VFC client name:fcs1            VFC client DRC:U8205.E6B.10BBD5P-V3-C6

Name          Physloc                            ClntID ClntName       ClntOS
------------- ---------------------------------- ------ -------------- -------
vfchost2      U8205.E6B.10BBD5P-V2-C9                 4 AIX119057-HA   AIX

Status:LOGGED_IN
FC name:fcs1                    FC loc code:U78AA.001.WZSG5HN-P1-C5-T2
```

```
Ports logged in:3
Flags:a<LOGGED_IN,STRIP_MERGE>
VFC client name:fcs0            VFC client DRC:U8205.E6B.10BBD5P-V4-C9

Name          Physloc                          ClntID ClntName       ClntOS
------------- -------------------------------- ------ -------------- -------
vfchost3      U8205.E6B.10BBD5P-V2-C10              4 AIX119057-HA   AIX

Status:LOGGED_IN
FC name:fcs3                    FC loc code:U78AA.001.WZSG5HN-P1-C7-T2
Ports logged in:3
Flags:a<LOGGED_IN,STRIP_MERGE>
VFC client name:fcs1            VFC client DRC:U8205.E6B.10BBD5P-V4-C10
```

FC port `fcs1` and `fcs3` are mapping to node `aix119057` and `aix119058`.

```
VIO server 2
$ lsmap -all -npiv
Name          Physloc                          ClntID ClntName       ClntOS
------------- -------------------------------- ------ -------------- -------
vfchost7      U8205.E6B.10BBD4P-V2-C15              1 AIX119053-HA   AIX

Status:LOGGED_IN
FC name:fcs3                    FC loc code:U78AA.001.WZSGGNL-P1-C5-T2
Ports logged in:3
Flags:a<LOGGED_IN,STRIP_MERGE>
VFC client name:fcs0            VFC client DRC:U8205.E6B.10BBD4P-V1-C5

Name          Physloc                          ClntID ClntName       ClntOS
------------- -------------------------------- ------ -------------- -------
vfchost8      U8205.E6B.10BBD4P-V2-C16              1 AIX119053-HA   AIX

Status:LOGGED_IN
FC name:fcs5                    FC loc code:U78AA.001.WZSGGNL-P1-C7-T2
Ports logged in:3
Flags:a<LOGGED_IN,STRIP_MERGE>
VFC client name:fcs1            VFC client DRC:U8205.E6B.10BBD4P-V1-C6
```

FC port `fcs3` and `fcs5` are mapping to node `aix119057` and `aix119058`.

Besides the storage zone, a heartbeat zone must be created to contain FC port `fcs1&fcs3` from VIO server 1 and FC port `fcs3&fcs5` from VIO server2.

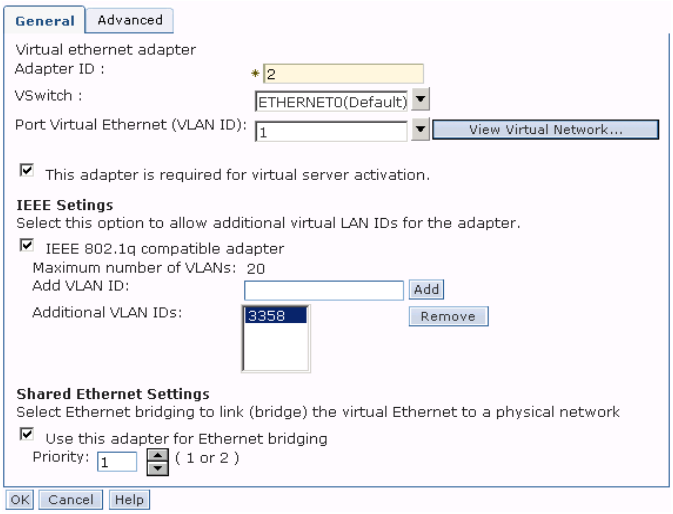
Next is to enable the target mode on these fcs device. All the following steps should perform on both VIO servers on each managed system.

```
VIO server 1
$ oem_setup_env
# chdev -l fscsi3 -a dyntrk=yes -a fc_err_recov=fast_fail
fscsi0 changed
# chdev -l fcs3 -a tme=yes
fscsi0 changed
# chdev -l fscsi1 -a dyntrk=yes -a fc_err_recov=fast_fail
fscsi0 changed
# chdev -l fcs1 -a tme=yes
fscsi0 changed

VIO server 2
$ oem_setup_env
# chdev -l fscsi3 -a dyntrk=yes -a fc_err_recov=fast_fail
fscsi0 changed
# chdev -l fcs3 -a tme=yes
fscsi0 changed
# chdev -l fscsi5 -a dyntrk=yes -a fc_err_recov=fast_fail
fscsi0 changed
# chdev -l fcs5 -a tme=yes
```

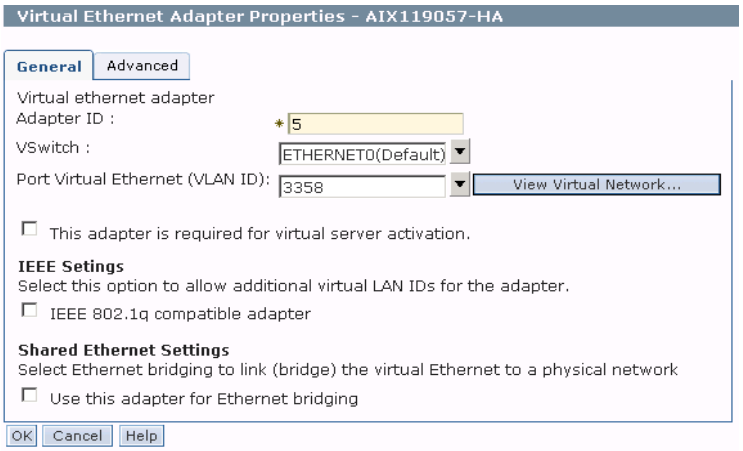
```
fscsi0 changed
```

At last a private VLAN 3358 need to be created.

Log in to each of the VIOS then add an additional VLAN to each shared Ethernet bridge adapter. This provides the VIOS connectivity to the VLAN 3358.



Respectively, create a virtual Ethernet adapter on each node, and set the port virtual VLAN ID to be 3358. This provides this node connectivity to the VLAN 3358.



Run the **`cfgmgr`** command to configure the newly added virtual Ethernet adapter.

Once this is complete, the SAN-based heartbeat is ready for use.

Tiebreaker disk

PowerHA 7.1.2 allows us to configure how the cluster will behave in the case that a partitioned or split condition occurs while we are defining the cluster.

A tiebreaker disk is used when nodes cannot communicate with each other. This communication failure results in the cluster splitting the nodes into two or more partitions. If failure occurs, both partitions attempt to lock the tie breaker disk. The partition that acquires the tie breaker disk continues to function, while the other partition reboots. Split and merge policy can be configured separately in SMIT menu.

It can be reached through **`smitty sysmirror`** **> Custom Cluster Configuration > Cluster Nodes and Networks > Initial Cluster Setup (Custom) > Configure Cluster Split and Merge Policy**.

```
Configure Cluster Split and Merge Policy for a Linked Cluster
Type or select values in entry fields.
Press Enter AFTER making all desired changes.
[Entry Fields]
Split Handling Policy None +
Merge Handling Policy Majority +
Split and Merge Action Plan Reboot +
Select Tie Breaker +
```

The split/merge policy attributes are described as follows:

- Cluster split policy

  The default behavior, called None, means that each partition becomes an independent cluster. This can be very dangerous because the resources will be activated in both partitions.

  The other option is tie breaker. The partition that survives is the one that is able to perform a SCSI-3 persistent reserve on the tiebreaker disk. All nodes in the losing partition will reboot.

- Cluster merge policy

  When nodes rejoin the cluster it is called a merge operation. What happens next depends on which split policy is chosen. Only the following two combinations are supported in current release when configure tiebreaker policy. That is, you have to use Majority if the split policy was None and TieBreaker if the split policy was Tie Breaker.

  – Split Policy: None,
  Merge Policy: Majority

  – Split Policy: TieBreaker
  Merge Policy: TieBreaker

A tiebreaker disk must be accessible by all nodes. However, the repository disk cannot be used as tie breaker again.

For example:
Site1: `aix119053`
Site2: `aix119057, aix119058`

Each node has five disks.

```
(0)aix119053/> lspv
hdisk0          00f6bbd447ed6f1b                    rootvg          active
hdisk1          00f6bbd4f347d521                    None
hdisk2          00f6bbd44b7f18cc                    None
hdisk3          00f6bbd44b7f1947                    None
hdisk4          00f6bbd4bf27c5ab                    None
hdisk5          00f6bbd4bf1ff5f0                    None

(0)aix119057/> lspv
hdisk0          00f6bbd54835a482                    rootvg          active
hdisk1          00f6bbd5bf8ba0d1                    None
hdisk2          00f6bbd5f34812a1                    None
hdisk3          00f6bbd54fdba0df                    None
hdisk4          00f6bbd4bf27c5ab                    None
hdisk5          00f6bbd4bf1ff5f0                    None
```

```
(0)aix119058/> lspv
hdisk0          00f6bbd5482c4b16                    rootvg          active
hdisk1         00f6bbd5bf8ba0d1                     None
hdisk2          00f6bbd54b7fdcea                    None
hdisk3          00f6bbd54fdba9de                    None
hdisk4          00f6bbd4bf27c5ab                    None
hdisk5          00f6bbd4bf1ff5f0                    None
```

The hdisk5 is accessible by all nodes. PowerHA uses SCSI-3 persistent reservation for handling site split/merge event, so next check whether hdisk5 is SCSI-3 persistent reserve-capable with the command **`lsattr -Rl hdiskX -a reserve_policy`**.

For example:

```
# lsattr -Rl hdisk5 -a reserve_policy
no_reserve
single_path
PR_exclusive
PR_shared
```

The **`PR_exclusive`** attribute is not supported by PowerPath 5.x so the powerpath pseudo device cannot be configured as a tiebreaker disk.

AIX LVM mirroring

The AIX LVM organizes and controls the system disk storage resources. Users and applications view logical volumes without regard to the actual physical components of the disks. The LVM, functions as a device driver layer above the regular device drivers to provide applications with a logical view of the system's disk storage. The LVM allows mirroring of up to three copies of a logical volume.

This section provides reference material for commands and utilities used in creating and managing host-level mirrors in AIX. Refer to the AIX documentation set for more detailed information on the AIX LVM. A general discussion of mirroring in the high-availability environment is provided in the section on mirroring.

You can use `SMIT` or standalone command line entries to set up or change AIX LVM logical volume mirroring. Define physical volumes using the methods outlined in Chapter 5, "AIX, VMAX/PowerMax Family Environment." Once you have hdisks defined, you create volume groups with multiple hdisks, and then logical volumes with multiple copies.

A clear understanding of the physical and logical device assignments of both the Symmetrix, VNX series, CLARiiON, and host systems is essential in planning a host-level mirroring policy. Consider both the physical disk location in the Symmetrix, VNX series, or CLARiiON system back end and the hardware path from the host when you define host-level mirrors. Plan the configuration of Symmetrix, VNX series, or CLARiiON and AIX volumes to ensure that mirror copies are on separate physical disks, and are accessed through separate channel paths. Keep in mind that physical volumes defined by AIX may in fact be on the same hyper-split, Symmetrix, VNX series, or CLARiiON device.

Although AIX LVM can automatically assign its physical partition copies to hdisks within the volume group, you may wish to select specific hdisks for each volume copy to be sure that the configuration is accomplished as designed.

Making mirrored volumes: Standalone commands

This section describes the steps required to create new logical volumes with multiple copies, or add copies to existing LVs. The command line entry method is presented below. Refer to "Making mirrored volumes using SMIT" on page 328 if you prefer to use `SMIT`.

Making a volume group

AIX permits you to group several physical disk devices into a volume group. Each AIX system can have 1 to 255 volume groups. A volume group is a collection of 1 to 32 physical volumes of varying size and type. A volume group that will contain mirrored copies should have at least one physical volume (**`hdisk`**) for each copy.

In the following examples, a volume group is defined with three physical devices. After you define the volume group, you can choose to create logical volumes with up to three copies; one copy on each device.

To make a volume group for Symmetrix devices, enter a statement similar to the following:

```
mkvg -s 8 -y testvg hdisk2 hdisk3 hdisk4
```

This statement creates a volume group named testvg from the physical disk devices, `hdisk2`, `hdisk3`, and `hdisk4`. The partition size of 8 MB (`-s 8`) can be used to define logical volumes up to approximately 8 GB. The default partition size is 4 MB.

When the **mkvg** command is executed, AIX creates the volume group comprising the specified physical volumes and activates the volume group using the **varyonvg** command.

Creating logical volumes with copies

Now that the volume groups are defined and varied on, you can define logical volumes with up to three copies.

The following example illustrates how to make the logical volume, `testlv`, in volume group `testvg`. `testlv` is a raw volume with 100 physical partitions, 800 MB in size. One copy of `testlv` is placed on hdisk2, and a second copy is on hdisk3.

```
mklv -y testlv -c 2 testvg 100 hdisk2 hdisk3
```

AIX creates the logical volume with 100 logical partitions. Two sets of 100 physical partitions are mapped to the two hdisks specified. The default strict allocation policy for the **mklv** command ensures that the logical partition copies do not share the same physical volume.

The logical volumes can be used as raw devices, or file systems can be added using the same procedures used for non-mirrored volumes.

Adding copies to a logical volume

You can add copies of an existing logical volume using the **mklvcopy** command. Make sure that the volume group has the physical volumes that will hold the LV copy. For example, to view the hdisks used for logical volume `testlv`, type:

```
lsvg -l testvg
```

Additional hdisks can be added to the volume group using the **extendvg** command. For example, to add physical volumes hdisk3 and hdisk4 to the `testvg` volume group, type:

```
extendvg testvg hdisk3 hdsk4
```

Once you have the correct physical volumes defined as a hdisks in the volume group, add a copy of an existing volume using a command similar to the following:

```
mklvcopy testlv 3 hdisk3 hdisk4
```

A copy of the logical volume testlv (previously defined on `hdisk2`) is now on `hdisk2`, `hdisk3`, and `hdisk4`. You can verify this by using `lslv testlv` to display the characteristics of the logical volume `testlv`. Each hdisk should now contain 100 PPs (the size of the original `testlv` volume).

The new copies will be synchronized to the original logical volume when the volume group is varied on. You can also use the **syncvg** command to initiate synchronization. If you wish to synchronize your copies as soon as they are created, include the **-k** option in the **mklvcopy** command line.

Making mirrored volumes using SMIT

The previous sections added logical volume copies using standalone AIX commands. These host-level mirrors can also be added using the `SMIT` menu.

To add mirrored logical volumes using `SMIT`, follow these steps:

1. Log in as **root**.
2. Launch `SMIT` by typing **smit** (if you are using an ASCII terminal) or **smitty** (if you are using an X terminal) and pressing **Enter**.

Adding a volume group

To create and name a volume group and select the physical volumes for that group:

1. Select **Physical and Logical Storage** from the `SMIT` main menu.
2. Select **Logical Volume Manager** from the next menu.
3. Select **Volume Group** from the menu.
4. Select **Add a Volume Group** from the menu.
5. Enter a volume group name (for example, `testvg`) when prompted.
6. Tab down to the next field and enter the physical partition size for the disk. (The default size is 4 MB.)
7. Press **F4** and select a **Physical Volume Name** by placing a **>** next to the disks you want to select and pressing **F7**. Select a separate hdisk for each logical volume copy to be created.
8. Click **DO** or press **Enter**. If this action completes successfully, press **ESC** three times to return to the **Logical Volume Manager** menu.

Creating logical volumes with copies

To add logical volumes to the volume group:

1. Select **Logical Volumes** from the **Logical Volume Manager** menu, and then select **Add a Logical Volume**.
2. Press **F4** and select the volume group to which you are adding logical volume(s) from the list of existing volume groups. Place a **>** next to the volume group you want to select and press **F7**.
3. If you are adding the logical volume to a new volume group, enter a name for the volume group.
4. Enter a name for the logical volume when prompted.
5. Select the volume group name. This is the volume group to contain the logical volume. Pressing **F4** provides a pop-up list of available volume groups.
6. Enter the number of physical partitions.
7. Enter the physical volume name. Pressing **F4** provides a pop-up list of available volume groups. Use **F7** to select hdisks; one for each logical volume copy.
8. Toggle the number of copies of each logical partition field to **2** or **3**.
9. Change any other parameters as desired, such as data allocation policy.

   Do not change the strict allocation policy default (**yes**) set in the `Allocate each logical partition copy on a separate physical volume` field. A **yes** ensures that copies are created on separate physical volumes.
10. Click **DO** or press **Enter**.
11. Proceed to "Adding copies," next.

Adding copies

You can add copies of an existing logical volume using `SMIT` as follows:

1. Log in as **`root`**.
2. Launch `SMIT` using the `SMIT` fast path **`smit lvsc`** to access the **Set Characteristics of a Logical Volume** menu.
3. Select **Add a Copy to a Logical Volume** from the `SMIT` menu.

4. Select the logical volume you want to mirror. Pressing **F4** provides a pop-up list of available volume groups. Use **F7** to select hdisks; one for each logical volume copy. Press **Enter** to record your selection.
5. Use the **Tab** key to toggle the **NEW TOTAL** field to **2** or **3**, and press **Enter**. (This setting determines the number of copies of the original logical volume.)
6. At the **PHYSICAL VOLUME** field, list the physical volume names (**F4**) and select (**F7**) the specific `hdisks` to contain the mirror copy. The default setting uses all available physical volumes for the mirror copy.
7. At the **SEPARATE** field, toggle to **yes** to ensure that copies are made on separate physical volumes.
8. At the **SYNCHRONIZE** the data in the new logical partitions copies? field, select **Yes** to synchronize the copies immediately.
9. Press **Enter** to enable the settings selected.

PowerHA and PowerPath error notification

AIX error notification allows you to run your own shell scripts or programs automatically in response to specified errors appearing in the AIX error log. These can be hardware and software operator error messages that are logged in the AIX error log. Each time an error is logged in the system error log, the error notification daemon checks the defined notification objects. If the error log entry matches the selection criteria it runs the notify methods.

The notification objects are error conditions that includes a wide range of criteria to define. AIX already contains some predefined notification objects. This section explains how to add a PowerPath notification object to detect an "all path down condition."

This section includes the following information:

Defining custom automatic error notification

Before you configure the automatic error notification, you must have a valid PowerHA topology and resource configuration.

Automatic error notification should be configured only when the cluster is not running. The following entries will need to be populated with the correct values, as shown in Table 14, "Values," on page 331.

Table 14 Values (page 1 of 3)

Entry	Description
Notification Object Name	Enter a user-defined name that identifies the error notification object.
Persist across system restart?	Set this field to **Yes** if you want to use this notification object persistently. Set this field to **No** if you want to use this object until the next reboot.
Process ID for use by Notify Method	You can specify a process ID for the notify method to use or you can leave it blank. Objects that have a process ID specified should have the Persist across system restart field set to **No**.

Table 14 Values (page 2 of 3)

Entry	Description
Select Error Class	Choose the appropriate error class. Valid values are: • **None**: No error class match • **All**: All error classes • **Hardware**: Hardware error class • **Software**: Software error class • **Errlogger**: Operator notifications, messages from the **errlogger** command
Select error type	Identify the severity of error log entries. Valid values are: • **None**: No entry type to match • **All**: Match all error types • **PEND**: Impending loss of availability • **PERM**: Permanent • **PERF**: Performance degradation • **TEMP**: Temporary • **UNKN**: Unknown error type
Match Alertable Errors?	This field is provided for use by network management applications alert agents. Chose **None** to ignore this entry. Valid values are: • **All**: Match all alertable errors • **TRUE**: Matches alertable errors • **FALSE**:Matches non-alertable errors
Select Error Label	Select an error label associated with the accurate error identifier from the /usr/include/sys/errids.h file. Press **F4** for a listing. If your application supports the AIX system error log, you can specify the application-specific error label.
Resource Name	The name of the failing resource. For the hardware error class, this is the device name. For the software class, this is the name of the failing executable. Specify **All** to match all error labels.
Resource Class	For the hardware error class, the resource class is the device class. It is not applicable for software errors. Specify **All** to match all resources classes.
Resource Type	Enter the device type by which a resource is known in devices object for hardware error class. Specify **All** to match all resource types.

Table 14 Values (page 3 of 3)

Entry	Description
Notify Method	Enter the full-path name of the executable file, shell script, or command to be ran whenever an error is logged that matches the defined criteria. You can pass the following variables to the executable: $1 Error log sequence number $2 Error ID $3 Error class $4 Error type $5 Alert flag $6 Resource name of the failing device $7 Resource type of the failing device $8 Resource class of the failing device $9 Error log entry label

Configuring PowerPath custom automatic error notification

To configure a PowerPath custom automatic error notification, perform the following steps on each node in the cluster:

1. Be sure the cluster is not running.
2. Start system management for PowerHA: **`smitty hacmp`**
3. Select **Problem Determination Tools** > **HACMP Error Notification** > **Add a Notify Method**.

   The **Add a Notify Method** dialog box displays. An example is shown in Figure 16. Values have already been added.

```
                        Add a Notify Method

Type or select values in entry fields.
Press Enter AFTER making all desired changes.


* Notification Object Name              EMC_ALL_PATHS
* Persist across system restart?        Yes
  Process ID for use by Notify Method   0
  Select Error Class                    Hardware
  Select Error Type                     PERM
  Match Alertable errors?               All
  Select Error Label                    EMCP_ALL_PATHS_DEAD
* Resource Name                         All
* Resource Class                        All
* Resource Type                         All
* Notify Method                         /usr/es/sbin/cluster/diag/pp_cl_failover $6 $9
```

Figure 16 Add a Notify Method dialog box example

4. Type or select the values in the entry fields and press **Enter**.

Creating a script

Once the custom PowerPath notification is added, you will need to create the `pp_cl_failover` script to include any specific actions you would like the cluster to take. That script is created in the following directory: `/usr/es/sbin/cluster/diag`.

This example creates a notification script to message the users and halt the node. Create the script to include the following entries and make sure the permissions are set to 500.

```
#!/bin/ksh
# Syntax:
# pp_cl_failover [ resource_name ] [error_log_label]
DEVICE_NAME=$1
ERROR_LABEL=$2

PATH="$($(dirname ${0})/../utilities/cl_get_path all)"
export PATH

PROGNAME=$(basename ${0})
HA_DIR="$(cl_get_path)"

MSG="HACMP errnotify event ERROR: $ERROR_LABEL has occurred on device: $DEVICE_NAME.\n"

#
# Check if the user changed the location of the file from the default
# /tmp directory
#
HA_LOG_DIR="/var/hacmp/log"

dir_in_odm=`odmget -q"name=hacmp.out" HACMPlogs | grep value | awk '{print $3}' | sed 's/"//g'`

if [[ "X$dir_in_odm" != "X/var/hacmp/log" ]]
then
    HA_LOG_DIR=$dir_in_odm
fi

#
# Log the message to stdout as well as hacmp.out
#
cl_log 6000 "$MSG" $ERROR_LABEL $DEVICE_NAME
cl_echo 6000 "$MSG" $ERROR_LABEL $DEVICE_NAME 2>> $HA_LOG_DIR/hacmp.out

#
# Mail root
#
DATE=`/usr/bin/date`

echo $DATE > /tmp/notify.$$
echo $MSG >> /tmp/notify.$$
mail root < /tmp/notify.$$
/bin/rm -rf /tmp/notify.$$

# If you wish to add additional functionality to this script
# Please place it between the "####" lines

#### Start of user code
wall " SYSTEM BEING HALTED BECAUSE OF CATASTROPHIC EVENT"
sleep 5
/usr/sbin/halt -q
#### End of user code
exit 0
```

Testing error notification objects

PowerHA can emulate error log entries. To test the notification objects:

1. Start system management for PowerHA: **`smitty hacmp`**
2. Select **Problem Determination Tools > HACMP Error Notification > Emulate Error Log Entry**.
3. Select the following notification object from the list.

   `EMCP_ALL_PATHS_DEAD  EMCP_ALL_PATHS`
4. Check the error label, notification object name, and notify method.
5. Press **Enter** to run the emulation.

The `/usr/es/sbin/cluster/diag/`**`pp_cl_failover command`** creates the appropriate error log entry in the system error log and the errdemon command automatically starts the notify method.

Additional error notification objects

The following PowerPath error notification objects can also be used in defining additional notifications, as shown in Table 15:

Table 15 PowerPath error notification objects

Object Name	Error Type	Error Class	Description
EMCP_NA_PATHS_DEAD	INFO	H	FUNCTION DEGRADED
EMCP_ALL_PATHS_DEAD	PERM	H	UNABLE TO COMMUNICATE WITH DEVICE
EMCP_VOL_DEAD	PERM	H	PHYSICAL VOLUME DEFINED AS MISSING
EMCP_BUS_DEAD	INFO	H	BACK-UP PATH INOPERATIVE
EMCP_PATH_DEAD	PERM	H	CONNECTION FAILURE
EMCP_PATH_ALIVE	INFO	H	BACK-UP PATH STATUS CHANGE
EMCP_VOL_ALIVE	INFO	H	PHYSICAL VOLUME IS NOW ACTIVE
EMCP_VOL_RESTORED	INFO	H	DEVICE ACCESS PATH RESTORED

GPFS clusters

IBM General Parallel File System (GPFS) is a scalable high performance file Management infrastructure for AIX, Linux and Windows operating systems.

GPFS is designed to:

- Enable single global namespace across platforms
- Enable high performance common storage
- Eliminate copies of data
- Allow efficient use of storage.
- Simplify file management

GPFS leverages enhanced availability provided by VMAX and Symmetrix storage. These enhancements take advantage of the exploitation of GPFS Cluster Services Tiebreaker Disk, as well as SCSI-3 Persistent Reserve (PR) type code 0x7. SCSI-3 PR provides fast failover and recovery by virtually instantaneously fencing any host that is deemed not healthy.

This section contains the following information:

High Availability

GPFS uses a cluster mechanism called quorum to maintain data consistency in the event of a node failure.

Quorum operates on the principle of majority rule. This means that a majority of the nodes in the cluster must be successfully communicating before any node can mount and access a file system. The quorum keeps any nodes that are cut off from the cluster (for example, by a network failure) from writing data to the GPFS file system.

During node failure situations, quorum must be maintained so that the cluster can remain online. If quorum is not maintained because of node failure, GPFS unmounts local file systems on the remaining nodes and attempts to reestablish quorum, at which point file system recovery occurs. For this reason, when planning to deploy a GPFS cluster, carefully consider the number of quorum nodes.

Node quorum and node quorum with tiebreaker disks

GPFS can use any one of the following methods for determining quorum:

- Node quorum: This algorithm is the default for GPFS. Note the following information about node quorum:
  - Quorum is defined as one plus half of the explicitly defined quorum nodes in the GPFS cluster.
  - There are no default quorum nodes; you must specify which nodes have this role.
- Node quorum with tiebreaker disks: When running on small GPFS clusters, you might want to have the cluster remain online with only one surviving node. To achieve this, add a tiebreaker disk to the quorum configuration. Note the following information about this algorithm:
  - Allows you to run with as little as one quorum node available if you have access to a majority of the quorum disks.
  - Enabling these disks starts by designating one or more nodes as quorum nodes. Then, define one to three disks as tiebreaker disks in the GPFS cluster configuration by using the tiebreakerDisks parameter with the **`mmchconfig`** command.
  - Designate any disk in the file system to be a tiebreaker.

    Note: Although you may have one, two, or three tiebreaker disks, a good practice is to use an odd number of tiebreaker disks.

Among the quorum node groups that appear after an interconnect failure, only those having access to a majority of tiebreaker disks can be candidates to be in the survivor group. Tiebreaker disks must be directly attached to all quorum nodes. In other words, GPFS Network Shared Disks (NSDs) accessed over TCP/IP are not valid to be used as tiebreaker disk from those quorum nodes.

SCSI-3 Persistent Reserve

SCSI-3 Persistent Reserve provides a mechanism for reducing recovery times from node failures.

For a device to properly offer SCSI-3 Persistent Reservation support for GPFS, it must support SCSI-3 PERSISTENT RESERVE IN with a service action of REPORT CAPABILITIES. The REPORT CAPABILITIES must indicate support for a reservation type of Write Exclusive All Registrants.

- Persistent Reserve: To enable PR and to obtain recovery performance improvements when AIX MPIO or EMC PowerPath multipathing is deployed, the GPFS cluster requires a specific environment:
  - All nodes in the cluster must be running AIX.
  - All NSD server nodes must be running AIX.
  - All disks configured within the cluster must be EMC Symmetrix VMAX.
  - All disks configured within the cluster must be PR enabled and using device reserve policy "reserve_policy=PR_shared". Mixing of PR and NON-PR devices isn't supported at this time.
  - AIX MPIO multipathing deployment requires minimum EMC Enginuity version 5874.207.166
  - PowerPath multipathing deployment requires minimum PowerPath version `5.5.P04.b003.tar.gz` and Enginuity version `5876.82.57`.
  - PowerPath deployment requires the use of /var/mmfs/etc/nsddevices file. This file is distributed in the EMC ODM software package. Refer to "Implementing PowerPath Persistent Reserve" on page 340 for further detail.

You must explicitly enable PR in the GPFS cluster configuration by using the *usePersistentReserve* option with the **`mmchconfig`** command. If you set the option as follows, GPFS attempts to set up PR on all PR capable disks:

```
mmchconfig usePersistentReserve=yes
```

Note: You will need to shut down the GPFS cluster before the above configuration change and bring it back online for PR to be enabled.

Symmetrix multipathing and ODM device support

GPFS can be connected to Symmetrix VMAX Series storage arrays by using either Fibre Channel (FCP) or Fibre Channel over Ethernet (FCoE) protocols. Multipath I/O solutions support includes both AIX MPIO and EMC PowerPath.

AIX ODM with MPIO

AIX GPFS clusters using Symmetrix devices under the control of AIX MPIO require the installation of ODM filesets **`EMC.Symmetrix.aix.rte`** and **`EMC.Symmetrix.fcp.MPIO.rte`** on all nodes connected to the SAN. You can use the following command on the nodes to determine currently installed version.

```
aix039051/>lslpp -L EMC.Symm*

Fileset                    Level  State  Type  Description
----------------------------------------------------------------------------
EMC.Symmetrix.aix.rte       5.3.0.6   C    F  EMC Symmetrix AIX Support Software
EMC.Symmetrix.fcp.MPIO.rte  5.3.0.6   C    F   EMC Symmetrix FCP MPIO Support Software

Minimum EMC ODM Support Software versions:

   "EMC.AIX.6.0.0.0 - for hosts running AIX 7.1
   "EMC.AIX.5.3.0.6 - for hosts  running AIX 6.1
```

ODM device attribute support

MPIO devices that will be used for GPFS NSDs are:

- reserve_policy = no_reserve
  - Initial reserve policy setting for NSDs
- **reserve_policy = PR_shared**
  - NSDs when GPFS SCSI-3 Persistent Reserve is enabled

AIX ODM with PowerPath

AIX GPFS clusters using Symmetrix devices under the control of PowerPath require the installation of ODM filesets **`EMC.Symmetrix.aix.rte`** and **`EMC.Symmetrix.fcp.rte`** on all nodes connected to the SAN. You can use the following command on the nodes to determine currently installed version.

```
aix039051/>lslpp -L EMC.Symm*

Fileset                 Level  State  Type  Description
----------------------------------------------------------------------------
EMC.Symmetrix.aix.rte   5.3.0.6  C     F    EMC Symmetrix AIX Support Software
EMC.Symmetrix.fcp.rte   5.3.0.6  C     F    EMC Symmetrix FCP Support Software
```

The minimum ODM Support Software versions are:

- EMC.AIX.6.0.0.0 - for hosts running AIX 7.1
- EMC.AIX.5.3.0.6 - for hosts running AIX 6.1

ODM device attribute support with PowerPath version 5.3.1 P01

PowerPath with Symmetrix devices is supported with minimum PowerPath version 5.3.1 P01 when the volumes are configured with the following ODM device attributes:

For the hdiskpower devices that will be used for creating GPFS NSDs:

```
reserve_lock=no
```

Note: SCSI-3 Persistent Reserve is not supported with this version of PowerPath.

ODM device attribute support with PowerPath version 5.5 P0x

PowerPath with Symmetrix devices is supported with minimum PowerPath version 5.5.P04 when the Dell EMC volumes are configured with the following ODM device attributes.

For the `hdiskpower` devices that will be used for creating GPFS NSDs:

- **`reserve_policy = no_reserve`**
  - Initial reserve policy setting for NSDs
- `reserve_policy = PR_shared`
  - NSDs when GPFS SCSI-3 Persistent Reserve is enabled

Enabling EMC Symmetrix for GPFS

EMC Symmetrix storage enablement requirements for GPFS include:

- The requirement for the Symmetrix FA and device flag enablement is critical to the cluster configuration. There are two approaches to achieve this:
  - Enable the flags on the FA ports in the impl.bin file
    - Typically performed by a Dell EMC Service Representative
  - Enable the flags through Solutions Enabler for each initiator in the GPFS cluster.
    - This enablement should be carefully planed considering some of these changes are an offline and/or online process.

  Consult with your GPFS administrator for the best approach within your environment.

  For Enginuity 5874/5875/5876:

C	Disable this flag
SC3	Default Enabled
SPC2	Default Enabled
OS2007	Default Enabled
PP	Default Enabled
UWN	Default Enabled
ACLX	Default Enabled
EAN	Default Enabled

- When implementing SCSI-3 Persistent Reserve, the Persistent Reserve flag needs to be enabled for each Symmetrix logical volume.

  EMC impl.bin VolumeEdit logical volume flag:

SCSI persistent	Enable this flag

- When implementing changes through Solutions Enabler, the SCSI-3 Persistent Reserve flag should be enabled for each Symmetrix logical volume with the following attribute setting.

```
attribute=SCSI3_persistent_reserv;
```

Implementing PowerPath Persistent Reserve

The following sections demonstrate the enablement of SCSI-3 Persistent Reserve using a test configuration comprising of four-node GPFS cluster, as shown in Figure 17 on page 341.

- Nodes (`aix039051`, `aix039052`) are NSD servers that are SAN- attached to Symmetrix.
- Node (`aix039053`) is local SAN-attached to Symmetrix.
- Node (`aix039054`) is network-attached through a private 10G Ethernet network to both NSD servers.

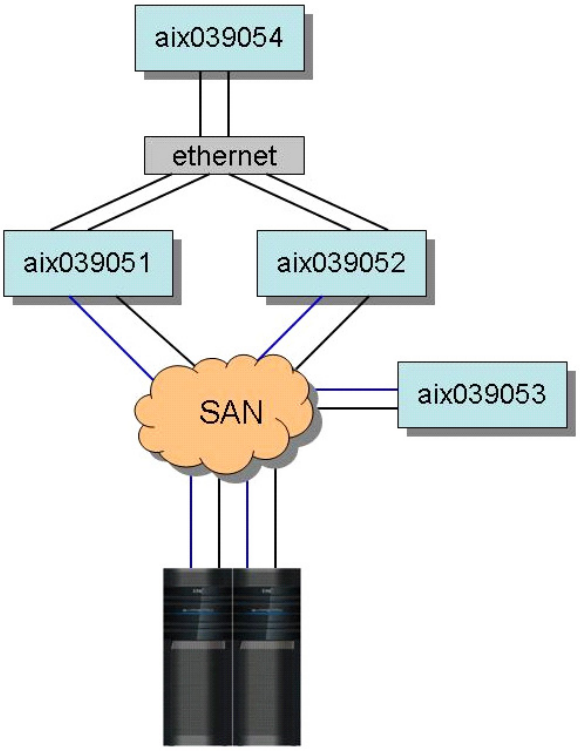


Figure 17 Four-node GPFS cluster example

GPFS NSD devices file requirements

The GPFS **`mmdevdiscover`** and **`nsddevices`** scripts are invoked by the **`mmfsd`** daemon when it tries to discover or verify physical devices previously defined in the GPFS cluster configuration with the **`mmcrnsd`** command. These scripts identify devices found in the `/dev` file system on the local node that may correlate to NSDs defined in GPFS cluster configuration.

GPFS uses the list of devices output by these scripts in mapping the NSD name listed in the configuration database to a corresponding local device in the `/dev` file system. When an NSD is created via the **`mmcrnsd`** command it is marked with a unique identifier written to sector two of the device. This unique identifier is recorded in the configuration database along with the user recognizable NSD name.

During GPFS disk discovery each device name output by **`mmdevdiscover`** and **`nsddevices`** is opened in turn and sector two of each device is read. If a match between an NSD identifier on the device and an identifier recorded in the configuration database is found, then this machine has local access to the NSD device.

In order for Dell EMC `hdiskpower` devices to be configured to support SCSI-3 Persistent Reserve, the GPFS `/var/mmfs/etc/nsddevices` file is required to correlate the *`DeviceType`* equivalency as an hdisk.

The *nsddevices* script is distributed in the ODM Support Software and located in */usr/lpp/EMC/Symmetrix/bin* directory. The following procedure should be used for proper installation.

1. Verify that the minimum ODM filesets `EMC.Symmetrix.aix.rte` and `EMC.Symmetrix.fcp.rte` are installed on all GPFS nodes connected to the SAN. You can use the following command on a node to determine the current installed version.

```
aix039051/>lslpp -L EMC.Symm*

Fileset                 Level  State  Type  Description
----------------------------------------------------------------------------
EMC.Symmetrix.aix.rte  5.3.0.6  C     F    EMC Symmetrix AIX Support Software
EMC.Symmetrix.fcp.rte  5.3.0.6  C     F    EMC Symmetrix FCP Support Software
```

The minimum ODM Support Software versions are:

- EMC.AIX.6.0.0.0 - for hosts running AIX 7.1
- EMC.AIX.5.3.0.6 - for hosts running AIX 6.1

2. Copy the **nsddevices** file into the appropriate directory.

```
aix039051/>cp /usr/lpp/EMC/Symmetrix/bin/nsddevices /var/mmfs/etc
```

3. Change the permissions on the file.

```
aix039051/>chmod +x /var/mmfs/etc/nsddevices
```

Enabling GPFS cluster SCSI-3 Persistent Reserve

To enable cluster support for SCSI-3 Persistent Reserve, the administrator must enable the `usePersistentReserve` attribute for the cluster.

For example, this test cluster queried the GPFS cluster configuration by issuing the **mmlsconfig** command.

```
aix039051/>mmlsconfig
Configuration data for cluster aix039051.lss.emc.com:
-----------------------------------------------------
clusterName aix039051.lss.emc.com
clusterId 789861885974439819
autoload no
minReleaseLevel 3.3.0.2
dmapiFileHandleSize 32
maxFilesToCache 3000
maxStatCache 12000
[aix039051,aix039052,aix039054]
subnets 10.0.10.0
[common]
failureDetectionTime 10
[aix039051]
psspVsd no
[common]
adminMode allToAll
```

The output shows that `usePersistentReserve` is not enabled because it is not included in the configuration list.

To enable `usePersistentReserve`, complete the following steps:

1. Verify the GPFS cluster is stopped on all nodes by issuing the following command:

```
aix039051/>mmshutdown -a
```

```
Node number  Node name        GPFS state
-----------------------------------------
       1      aix039051        down
       2      aix039052        down
       3      aix039053        down
       4      aix039054        down
```

2. Issue the following command to enable `usePersistentReserve`.

   ```
   aix039051/>mmchconfig usePersistentReserve=yes
   ```

3. Verify that it is now enabled for the cluster configuration:

```
aix039051/>mmlsconfig
Configuration data for cluster aix039051.lss.emc.com:
-----------------------------------------------------
clusterName aix039051.lss.emc.com
clusterId 789861885974439819
autoload no
minReleaseLevel 3.3.0.2
dmapiFileHandleSize 32
usePersistentReserve yes
maxFilesToCache 3000
maxStatCache 12000
[aix039051,aix039052,aix039054]
subnets 10.0.10.0
[common]
failureDetectionTime 10
[aix039051]
psspVsd no
[common]
adminMode allToAll
```

4. Start GPFS on all the cluster nodes using:

   ```
   aix039051/>mmstartup -a
   ```

   SCSI-3 Persistent Reserve is now enabled and applied to the active GPFS cluster configuration. The administrator can proceed in configuring the NSD devices and GPFS filesystem.

5. Verify that NSD devices file is created with your desired configuration options.

```
aix039051/>pg /usr/lpp/install/GPFS/gpfs.power.disks

hdiskpower0:aix039051:aix039052:::nsdpower0
hdiskpower1:aix039051:aix039052:::nsdpower1
hdiskpower2:aix039051:aix039052:::nsdpower2
hdiskpower3:aix039051:aix039052:::nsdpower3
hdiskpower4:aix039051:aix039052:::nsdpower4
hdiskpower5:aix039051:aix039052:::nsdpower5
```

6. Create the NSD devices.

```
aix039051/>mmcrnsd -F /usr/lpp/install/GPFS/gpfs.power.disks -v no

mmcrnsd: Processing disk hdiskpower0
mmcrnsd: Processing disk hdiskpower1
mmcrnsd: Processing disk hdiskpower2
mmcrnsd: Processing disk hdiskpower3
mmcrnsd: Processing disk hdiskpower4
mmcrnsd: Processing disk hdiskpower5
mmcrnsd: 6027-1371 Propagating the cluster configuration data to all
affected nodes.  This is an asynchronous process.
```

7. Verify the devices file was updated after the creation of the NSD devices.

```
aix039051/>pg /usr/lpp/install/GPFS/gpfs.power.disks

# hdiskpower0:aix039051:aix039052:::nsdpower0
nsdpower0:::dataAndMetadata:4001::
# hdiskpower1:aix039051:aix039052:::nsdpower1
nsdpower1:::dataAndMetadata:4001::
# hdiskpower2:aix039051:aix039052:::nsdpower2
nsdpower2:::dataAndMetadata:4001::
# hdiskpower3:aix039051:aix039052:::nsdpower3
nsdpower3:::dataAndMetadata:4001::
# hdiskpower4:aix039051:aix039052:::nsdpower4
nsdpower4:::dataAndMetadata:4001::
# hdiskpower5:aix039051:aix039052:::nsdpower5
nsdpower5:::dataAndMetadata:4001::
```

8. Verify the NSD devices you just created are part of the free list.

```
aix039051/>mmlsnsd

 File system   Disk name    NSD servers
---------------------------------------------------------------------------
 (free disk)   nsdpower0    aix039051.lss.emc.com,aix039052.lss.emc.com
 (free disk)   nsdpower1    aix039051.lss.emc.com,aix039052.lss.emc.com
 (free disk)   nsdpower2    aix039051.lss.emc.com,aix039052.lss.emc.com
 (free disk)   nsdpower3    aix039051.lss.emc.com,aix039052.lss.emc.com
 (free disk)   nsdpower4    aix039051.lss.emc.com,aix039052.lss.emc.com
 (free disk)   nsdpower5    aix039051.lss.emc.com,aix039052.lss.emc.com
```

As part of the process to enable SCSI-3 PR in the cluster configuration, GPFS will generate and distribute a set of PR keys unique to each host in the cluster and to the Storage device. Once PR is successfully enabled, all the hosts, storage devices, and FAs that are provisioned to the shared disks will be made aware of the PR keys that are in use. Since PR key's uniqueness is per-host, the same PR key will be used on each path, even if there are multiple paths to a single device.

IMPORTANT

In some cases the propagation of the PR keys may not have been enabled for some devices that are SAN-attached. Enablement completion will actually take place once the GPFS filesystem is created and mounted for the first time. All subsequent cluster startup and filesystem mounts will not be affected by this behavior.

Note: The creation of the NSD devices will take longer than a MPIO configuration because the PR keys needs to be processed for the `hdiskpower` device and all paths to its hdisk devices. In configurations where there are a large number of hdiskpower disks and paths in use, the configuration setup can take an extended amount of time.

Dell EMC provides the option of setting up the PR keys prior to creating the NSD devices. This can be achieved by running the Dell EMC `nsdprsetup.sh` script. The preferred way is to have GPFS perform this function considering it executes NSD disk verification that the nsdprsetup.sh script does not perform because the NSD devices have not yet been configured.

Refer to "Implementing PowerPath reference" on page 346 for further details.

9. Verify PR has been enabled on the NSD Server devices.

```
aix039051/>mmlsnsd -X
```

```
 Disk name  NSD volume ID        Device Devtype            Node name                  Remarks
-------------------------------------------------------------------------------------------------
 nsdpower0    0AF627334F8844E3   /dev/hdiskpower0 hdisk    aix039051.lss.emc.com    server node,pr=yes
 nsdpower0    0AF627334F8844E3   /dev/hdiskpower0 hdisk    aix039052.lss.emc.com    server node,pr=yes
 nsdpower1    0AF627334F884502   /dev/hdiskpower1 hdisk    aix039051.lss.emc.com    server node,pr=yes
 nsdpower1    0AF627334F884502   /dev/hdiskpower1 hdisk    aix039052.lss.emc.com    server node,pr=yes
 nsdpower2    0AF627334F884521   /dev/hdiskpower2 hdisk    aix039051.lss.emc.com    server node,pr=yes
 nsdpower2    0AF627334F884521   /dev/hdiskpower2 hdisk    aix039052.lss.emc.com    server node,pr=yes
 nsdpower3    0AF627334F884540   /dev/hdiskpower3 hdisk    aix039051.lss.emc.com    server node,pr=yes
 nsdpower3    0AF627334F884540   /dev/hdiskpower3 hdisk    aix039052.lss.emc.com    server node,pr=yes
 nsdpower4    0AF627334F88455F   /dev/hdiskpower4 hdisk    aix039051.lss.emc.com    server node,pr=yes
 nsdpower4    0AF627334F88455F   /dev/hdiskpower4 hdisk    aix039052.lss.emc.com    server node,pr=yes
 nsdpower5    0AF627334F88457E   /dev/hdiskpower5 hdisk    aix039051.lss.emc.com    server node,pr=yes
 nsdpower5    0AF627334F88457E   /dev/hdiskpower5 hdisk    aix039052.lss.emc.com    server node,pr=yes
```

You can query PR keys registered for a device by running the **tsprreadkeys** command. The following results were returned when this command was executed on our test cluster. As indicated, there are four duplicate keys for each host that is registered because each host hba initiator is mapped to two FA ports. The key for host aix039051=00006d000af62733 and aix039052=00006d000af62734. Also notice that the PR keys for the SAN-attached host aix039053 keys is not in the output. These keys will be enabled when the GPFS filesystem is mounted for the first time.

```
aix039051/>tsprreadkeys hdiskpower0

Registration keys for hdiskpower0
1. 00006d000af62733
2. 00006d000af62734
3. 00006d000af62733
4. 00006d000af62734
5. 00006d000af62733
6. 00006d000af62734
7. 00006d000af62733
8. 00006d000af62734
```

Creating and verifying the GPFS filesystem

Now that the NSD devices have been created, configure a GPFS filesystem using those NSDs with **mmcrfs** command.

```
aix039051/>mmcrfs /gpfs1 /dev/gpfs1 -F /usr/lpp/install/GPFS/gpfs.power.disks -B 1024k -N
2000000 -n 8 -A yes -v no

GPFS: 6027-531 The following disks of gpfs1 will be formatted on node aix039051:
    nsdpower0: size 5894400 KB
    nsdpower1: size 5894400 KB
    nsdpower2: size 5894400 KB
    nsdpower3: size 5894400 KB
    nsdpower4: size 5894400 KB
    nsdpower5: size 5894400 KB
GPFS: 6027-540 Formatting file system ...
GPFS: 6027-535 Disks up to size 49 GB can be added to storage pool 'system'.
Creating Inode File
Creating Allocation Maps
Clearing Inode Allocation Map
Clearing Block Allocation Map
Formatting Allocation Map for storage pool 'system'
GPFS: 6027-572 Completed creation of file system /dev/gpfs1.
mmcrfs: 6027-1371 Propagating the cluster configuration data to all
  affected nodes.  This is an asynchronous process.
```

Verify that the GPFS filesystem has been created using the predefined hdiskpower devices.

```
aix039051/> mmlsnsd

 File system   Disk name    NSD servers
---------------------------------------------------------------------------
```

```
 gpfs1         nsdpower0    aix039051.lss.emc.com,aix039052.lss.emc.com
 gpfs1         nsdpower1    aix039051.lss.emc.com,aix039052.lss.emc.com
 gpfs1         nsdpower2    aix039051.lss.emc.com,aix039052.lss.emc.com
 gpfs1         nsdpower3    aix039051.lss.emc.com,aix039052.lss.emc.com
 gpfs1         nsdpower4    aix039051.lss.emc.com,aix039052.lss.emc.com
 gpfs1         nsdpower5    aix039051.lss.emc.com,aix039052.lss.emc.com
```

GPFS determines if a node has physical or virtual connectivity to an underlying NSD disk through a sequence of commands invoked from the GPFS daemon. This determination is called *disk discovery* and occurs at both initial GPFS startup as well as whenever a file system is mounted.

The default order of access used in disk discovery is:

1. Local block device interfaces for SAN.
2. NSD servers.

The *useNSDserver* file system mount option can be used to set the order of access used in disk discovery. The values that can be used are:

always	Always access the disk using the NSD server.
as found	Access the disk as found.
as needed	Access the disk any way possible. *This is the default.*
never	Always use local disk access.

Mount the filesystem with one of the valid options:

```
aix039051/>mmmount gpfs1 -a -o useNSDserver=always
Mon Apr 16 09:51:07 CDT 2012: 6027-1623 mmmount: Mounting file systems ...
```

The GPFS SCSI-3 PR enablement is now completed. Run the following command in order to verify all PR keys are set up for all SAN-attached devices. Note that the keys for host `aix039053` are now listed in the output of the **`tsprreadkeys`** command.

```
aix039051/>tsprreadkeys hdiskpower0
Registration keys for hdiskpower0
1.  00006d000af62733
2.  00006d000af62734
3.  00006d000af62735
4.  00006d000af62733
5.  00006d000af62734
6.  00006d000af62735
7.  00006d000af62733
8.  00006d000af62734
9.  00006d000af62735
10. 00006d000af62733
11. 00006d000af62734
12. 00006d000af62735
```

Implementing PowerPath reference

In order for Dell EMC `hdiskpower` devices to be configured as SCSI-3 Persistent Reserve, the `/var/mmfs/etc/nsddevices` file is required to correlate the `DeviceType` equivalency as an `hdisk`.

```
# When properly installed, this script is invoked by the
# /usr/lpp/mmfs/bin/mmdevdiscover script.
#
```

```
# INSTALLATION GUIDELINES FOR THIS SCRIPT:
#
#   a) Install this script using the configuration guidelines below
#   b) copy this script to /var/mmfs/etc/nsddevices
#   c) ensure this script is executable (chmod +x /var/mmfs/etc/nsddevices)
#
# DESCRIPTION OF NSD DEVICE DISCOVERY:
#
#   The mmdevdiscover script and conversely this script are invoked by
#   the mmfsd daemon when it tries to discover or verify physical
#   devices previously defined to GPFS with the mmcrnsd command.  These
#   scripts identify devices found in the /dev file system on the local
#   machine that may correlate to NSDs defined to GPFS.
#
#   GPFS uses the list of devices output by these scripts in mapping
#   the NSD name listed in the configuration database to a local device
#   in the /dev file system.  When an NSD is created via the mmcrnsd
#   command it is marked with a unique identifier written to sector
#   two of the device.  This unique identifier is recorded in the
#   configuration database along with the user recognizable NSD name.
#
#   During GPFS disk discovery each device name output by mmdevdiscover
#   and nsddevices is opened in turn and sector two of each device is
#   read.   If a match between an NSD identifier on the device and an
#   identifier recorded in the configuration database is found, then
#   this machine has local access to the NSD device.  I/O is thus
#   subsequently performed via this local /dev interface.
#
# CONFIGURATION AND EDITING GUIDELINES:
#
#   If this script is not installed then disk discovery is done
#   only via the commands listed in mmdevdiscover.
#
#   The output from both this script and nsddevices is a number
#   of lines in the following format:
#
#     deviceName deviceType
#
#   where (deviceName) is a device name such as (hdiskpower1)
#   and (deviceType) is a set of known disk types as (hdisk).
#
#     /usr/lpp/mmfs/bin/mmdevdiscover
#
#   for a list of currently known deviceTypes
#
#   Example output:
#
#     hdiskpower1  hdisk
#     hdiskpower2  hdisk
#
# krichards 12/06/2010
# nsddevices.power v1.0.0.0
################################################################################
#set -x
osName=$(/bin/uname -s)
DISK=powerdisk
PPLIST=`lsdev -Cc disk -t power -SA -F "name"`

if [[ $osName = AIX ]]
then
        for PWR in $PPLIST;do
        echo "$PWR  hdisk"
done
fi
```

PowerPath nsdprsetup.sh script

The creation of `hdiskpower` NSD devices will take longer than a typical MPIO configuration because the PR keys need to be processed for the hdiskpower device and across all the paths to its hdisk devices. In configurations where there are a large number of `hdiskpower` disks and paths in use, the configuration setup can take an extended amount of time.

EMC provides the option of setting up the PR keys prior to creating the NSD devices. This can be achieved by running the Dell EMC `nsdprsetup.sh` script. However, the preferred method is to have GPFS perform this function considering it executes NSD disk verification which the `nsdprsetup.sh` script in not able to handle since the NSD devices have not yet been configured.

```
# When properly installed, this script is invoked by the user to create
# Persistent Reserve Key on PowerPath devices.
#
# /usr/lpp/EMC/Symmetrix/bin/nsdprsetup.sh script.
#
# INSTALLATION GUIDELINES FOR THIS SCRIPT:
#
#   a) Install this script using the configuration guidelines below
#   b) ensure this script is executable (chmod +x
#      /usr/lpp/EMC/Symmetrix/bin/nsdprsetup.sh)
#
# DESCRIPTION OF NSD PERSISTENT RESERVE KEY:
#
#   GPFS supports SCSI-3 PERSISTENT RESERVE WRITE EXCLUSIVE ALL REGISTRANTS at
#   version 3.2.1 and higher. Enabling device support can be automated through
#   GPFS command "mmchconfig userPersistentReserve=yes" but this process can take
#   a long time if there are many hdiskpower and paths in the configuration.
#   This script can be used to minimize the setup time.
#
#   The PR_key_value value is determined by the host EMCMagicNumber
#   and ip address in "hex".
#
#   EMCMagicNumber=0x6d00
#   Host ip address in dec=10.24.39.51 converted to hex=0af62733
#
#
# CONFIGURATION AND SETUP GUIDELINES:
#
#   In order for this script to execute correctly it requires a NSD DeviceFile for
#   input. The content of the file should look something like this:
#
#       hdiskpower0
#       hdiskpower1
#       hdiskpower2
#       hdiskpower3
#
#       NOTE: There can exist a difference in the number ordering of hdiskpower
#             devices across the nodes in the cluster. Verify the NSD DeviceFile
#             has the correct hdiskpower entries. Device verification can be
#             compared across the nodes using the command:
#             powermt display dev=hdiskpower#
#
#                          OR
#
#             /usr/lpp/EMC/Symmetrix/bin/inq.aix64_51
#
#   1. Create a NSD DeviceFile. This is the list of hdiskpower devices
#      which will be used.
#   2. Run the script in order to enable SCSI-3 Persistent Reserve
#   3. After completion of the script the devices should have the following
```

```
#       attributes enabled. PR key will differ across hosts:
#
#       PR_key_value   0x6d000af62733
#       reserve_policy PR_shared
#    4. When inputting your ip address make sure you use the address that
#       resolves your hostname. The key that will be generated uses a decimal
#       to Hex conversion.
#
#       (i.e) hostname = aix039051
#             ipdaddress = 10.246.39.51
#             key interpretation = EMCMagicNumber and ipdaddress hex conversion
#             generated key = 0x6d000af62733
#
# krichards 11/07/2011
# prsetup.sh v1.0.0.0
################################################################################
# Generate an 8 byte key:
#   bytes 0-1    x'6d00' EMCMagicNumber
#   bytes 2-5    hex ipaddr of node
#   bytes 7-8    reserved

EMCMagicNumber=6d00
osName=$(/bin/uname -s)

if [[ $osName = AIX ]]
then
        echo 'Input host IP Address to generate PRKeyValue'
        read ipaddress
        finaladdress=`echo $ipaddress|awk -F. '{print $1, $2, $3, $4}'`
        PRKeyValue=$(printf "0x%s%2.2x%2.2x%2.2x%2.2x" $EMCMagicNumber $finaladdress)
        echo 'Input complete path name to NSD device file'
        read nsdfile
        NSDLIST=`pg $nsdfile | grep hdiskpower|awk`
                for PWR in $NSDLIST;do
                echo ""
                chdev -l $PWR -a PR_key_value=$PRKeyValue > /dev/null 2>&1
                        if [ $? -eq 0 ] ; then
                        echo "$PWR PR_key_value=$PRKeyValue"
                        else
                        echo "$PWR PR_key_value change failed"
                        fi
                chdev -l $PWR -a reserve_policy=PR_shared > /dev/null 2>&1
                        if [ $? -eq 0 ] ; then
                        echo "$PWR reserve_policy=PR_shared"
                        else
                        echo "$PWR reserve_policy=PR_shared change failed"
                        fi
                echo ""
done
fi
```

GPFS mmdf command with PowerPath devices

The GPFS mmdf command is used to query the available file space on a GPFS cluster. For each disk in the GPFS file system, the **mmdf** command displays the following information, by failure group and storage pool:

- Size of the disk.
- Failure group of the disk.
- Whether the disk is used to hold data, metadata, or both.
- Available space in full blocks.
- Available space in fragments.

However, for hdiskpower devices, the **`mmdf`** command output does not include the information at an individual disk level as indicated below. The GPFS administrator needs to reference the total space available and in use.

```
aix039051/usr/lpp/install/GPFS>mmdf gpfs1
Skipping snapshot  (snapId 0).
Disks in storage pool: system (Maximum disk size allowed is 49 GB)
disk                disk size  failure holds    holds              free KB            free KB
name                    in KB    group metadata data        in full blocks       in fragments
--------------- ------------- -------- -------- ----- -------------------- -------------------
disk                disk size  failure holds    holds              free KB            free KB
name                    in KB    group metadata data        in full blocks       in fragments
--------------- ------------- -------- -------- ----- -------------------- -------------------
disk                disk size  failure holds    holds              free KB            free KB
name                    in KB    group metadata data        in full blocks       in fragments
--------------- ------------- -------- -------- ----- -------------------- -------------------
disk                disk size  failure holds    holds              free KB            free KB
name                    in KB    group metadata data        in full blocks       in fragments
--------------- ------------- -------- -------- ----- -------------------- -------------------
disk                disk size  failure holds    holds              free KB            free KB
name                    in KB    group metadata data        in full blocks       in fragments
--------------- ------------- -------- -------- ----- -------------------- -------------------
disk                disk size  failure holds    holds              free KB            free KB
name                    in KB    group metadata data        in full blocks       in fragments
--------------- ------------- -------- -------- ----- -------------------- -------------------
                -------------                         -------------------- -------------------
(pool total)         35366400                              35141632 ( 99%)        12288 ( 0%)

                =============                         ==================== ===================
(total)              35366400                              35141632 ( 99%)        12288 ( 0%)

                -------------                         -------------------- -------------------
(total)                     0                                     0 (  0%)            0 ( 0%)
           [--pnfs | --nopnfs]
```

Disabling SCSI-3 PR for EMC PowerPath devices within a GPFS cluster

GPFS and Persistent Reserve (PR) functionality is used to improve failover times. If PR is already enabled and there is a requirement to disable it for the cluster, running the command **`mmchconfig usePersistentReserve=no`** is not enough since the GPFS daemons are unable to process the disablement of PR for the `hdiskpower` devices properly.

The administrator must shutdown the cluster on all nodes, manually change the hdiskpower device attributes to *reserve_policy=no_reserve and PR_key_value=none* using the **`chdev`** command (that is, **`chdev –l hdiskpower0 –a "reserve_policy=no_reserve PR_key_value=none"`**). Once this is completed the administrator can run the **`mmchconfig usePersistentReserve=no`** command and start up the cluster.

PART 4

Virtualization

Part 5 includes:

CHAPTER 13

Dell EMC VPLEX

This chapter describes VPLEX-specific configuration in the IBM AIX environment and contains support information.

VPLEX overview

The VPLEX family is a solution for federating Dell EMC and non-Dell EMC storage. The VPLEX family is a hardware and software platform that resides between the servers and heterogeneous storage assets supporting a variety of arrays from various vendors. VPLEX can extend data over distance within, between, and across data centers. VPLEX simplifies storage management by allowing LUNs, provisioned from various arrays, to be managed through a centralized management interface.

For more details, refer to VPLEX documentation, available at Dell EMC Online Support. Refer to the following documents for configuration and administration operations:

- *EMC VPLEX with GeoSynchrony 5.0 Product Guide*
- *EMC VPLEX with GeoSynchrony 5.0 CLI Guide*
- *EMC VPLEX with GeoSynchrony 5.0 Configuration Guide*
- *EMC VPLEX Hardware Installation Guide*
- *EMC VPLEX Release Notes*
- *Implementation and Planning Best Practices for EMC VPLEX Technical Notes*
- *VPLEX online help, available on the Management Console GUI*
- *VPLEX Procedure Generator, available at Dell EMC Online Support*
- *Dell EMC Simple Support Matrices, EMC VPLEX and GeoSynchrony, available at Dell EMC E-Lab Navigator.*

For the most up-to-date support information, always refer to the Dell EMC Simple Support Matrices.

Prerequisites

Before configuring VPLEX in the IBM AIX environment, complete the following on each host:

- Confirm that all necessary remediation has been completed.
  This ensures that OS-specific patches and software on all hosts in the VPLEX environment are at supported levels according to the Dell EMC Simple Support Matrices.
- Confirm that each host is running VPLEX-supported failover software and has at least one available path to each VPLEX fabric.

  Note: Always refer to the Dell EMC Simple Support Matrices for the most up-to-date support information and prerequisites.

- *If a host is running PowerPath, confirm that the load-balancing and failover policy is set to **Adaptive**.*
- *Support requires minimum AIX 7100-00-02-1041, AIX 6100-04-03-1009, and AIX 5300-11-02-1009.*
- *AIX attach to VPLEX requires minimum EMC ODM 5.3.0.3.*
- *PowerPath, Veritas DMP, and AIX MPIO are the only multipath solutions supported with VPLEX and AIX hosts.*
- *AIX MPIO minimum requirements are as follows:*
  - Use ODM files on each host. This includes the reset_delay parameter set to 0. The ODM versions are posted at ftp://ftp.emc.com/pub/elab/aix/ODM_DEFINITIONS
    - AIX 6.1 Technology Level 6100-08-00-1241, 6100-07-06-1241 and later. Requires minimum ODM version 5.3.0.8
    - AIX 7.1 Technology Level 7100-02-00-1241, 7100-01-06-1241 and later. Requires minimum ODM version 6.0.0.3
    - VIOS 2.2.2.0 and later. Requires minimum ODM version 5.3.0.8

    The host must be rebooted after updating ODM file for the settings to take effect.
  - The following configurations with AIX 6.1 and 7.1 are supported and MPIO:
    - AIX LVM must be used
    - IBM PowerHA, GPFS filesystem
    - VIOS, VIOC, LPAR, NPIV
  - VPLEX 5.1.x is the minimum version supported.

IMPORTANT

For optimal performance in an application or database environment, ensure alignment of your host's operating system partitions to a 32 KB block boundary.

Veritas DMP settings with VPLEX

- *Veritas DMP 5.1 SP1 requires the asl package 5.1.100.100 to correctly detect the VPLEX array.*
- *If a host attached to VPLEX is running Veritas DMP multipathing, change the following values of the DMP tunable parameters on the host to improve the way DMP handles transient errors at the VPLEX array in certain failure scenarios:*
  - ***dmp_lun_retry_timeout*** for the VPLEX array to 60 seconds using the following command:

```
"vxdmpadm setattr enclosure emc-vplex0 dmp_lun_retry_timeout=60"
```

  - ***recoveryoption*** to throttle and ***iotimeout*** to 30 using the following command:

```
"vxdmpadm setattr enclosure emc-vplex0 recoveryoption=throttle iotimeout=30"
```

Configuring Fibre Channel HBAs with VPLEX

This section provides Fibre Channel HBA-related configuration details that must be addressed when using Fibre Channel with VPLEX.

The supported Fibre Channel HBA type is an IBM-branded Emulex adapter.

The values provided are required and optimal for most scenarios. However, in extreme scenarios the values may need to be tuned if the performance of the VPLEX shows high front-end latency in the absence of high back-end latency and this has visible impact on host application(s). This may be an indication that there are too many outstanding IOs at a given time per port. For further information on how to monitor VPLEX performance, refer to the "Performance and Monitoring" section of the *VPLEX Administration Guide*, located at Dell EMC Online Support. If the host application(s) is seeing a performance issue with the required settings, contact EMC Support for further recommendations.

This section includes the following information:

Overview

Note: Changing the queue depth is designed for advanced users. Increasing the queue depth may cause the AIX host to over-stress arrays connected to it, resulting in performance degradation while performing IO.

---

num_cmd_elems settings control the amount of outstanding I/O requests per HBA port. This can be done on the HBA firmware level using the **chdev** command.

Queue depth settings control the amount of outstanding I/O requests per disk/path (per ITL). This can be done using the **chdev** command.

The following procedures detail adjusting the queue depth and num_cmd_elems settings as follows:

- *Set the disk queue depth in AIX to its default value of 20 (decimal).*
- *Set the HBA adapter num_cmd_elems in the range of 32(decimal) to 64 (decimal) for larger IO block size or to its default value of 200 (decimal) for smaller IO block size.*

Note: This setting of disk queue depth means 20 outstanding IOs per ITL, so if a host has 4 paths then there are 20 outstanding IOs per path, resulting in a total of 80.

---

max_xfer_size attribute controls the maximum IO size the adapter device driver will handle, also controls a memory area used by the adapter for data transfers. The default value of max_xfer_size is 0x100000 (ó16 MB) and the maximum value is 0x1000000 (ó128 MB).

lg_term_dma attribute controls the direct memory access (DMA) memory resource that an adapter driver can use. The default value of lg_term_dma is 0x200000 (ó2 MB), and the maximum value is 0x8000000 (ó128 MB).

For larger IO Block size, along with queue depth and num_cmd_elems settings, the attributes max_xfer_size and lg_term_dma should be set to the value of 0x1000000 (ó128 MB) and 0x2000000 (ó32 MB) respectively.

For smaller IO Block size, along with queue depth and `num_cmd_elems` settings, the `max_xfer_size` and `lg_term_dma` attributes should be set to its default value of 0x100000 (ó16 MB) and 0x200000 (ó2 MB) respectively.

Setting queue depth on disk

To set the queue depth on the disk, complete the following steps.

1. To check the existing queue depth, run the **`lsattr`** command for each disk.

   `lsattr -l hdisk# -E`

   Note: Replace `hdisk#` with the volume number (`hdisk0`, `hdisk1`, etc.). Disk name depends on the Multipath being used. That is, the VPLEX volume presented by PowerPath to the AIX host is denoted as `hdiskpower#` (`hdiskpower0`, `hdiskpower1`, etc.).

```
# lsattr -l hdisk2 -E
PCM             PCM/friend/MINVISTA_DISK Path Control Module             True
PR_key_value    none                     Persistant Reserve Key Value    True
algorithm       round_robin              Algorithm                       True
clr_q           yes                      Device CLEARS its Queue on error True
dist_err_pcnt   0                        Distributed Error Percentage    True
dist_tw_width   50                       Distributed Error Sample Time   True
hcheck_cmd      test_unit_rdy            Health Check Command            True
hcheck_interval 60                       Health Check Interval           True
hcheck_mode     nonactive                Health Check Mode               True
location                                 Location Label                  True
lun_id          0x0                      Logical Unit Number ID          False
lun_reset_spt   yes                      FC Forced Open LUN              True
max_coalesce    0x100000                 Maximum Coalesce Size           True
max_retries     5                        Maximum Number of Retries       True
max_transfer    0x100000                 Maximum TRANSFER Size           True
node_name       0x5000144046e07ef4       FC Node Name                    False
pvid            none                     Physical volume identifier      False
q_err           no                       Use QERR bit                    True
q_type          simple                   Queue TYPE                      True
queue_depth     32                       Queue DEPTH                     True
reassign_to     120                      REASSIGN time out value         True
reserve_policy  no_reserve               Reserve Policy                  True
reset_delay     0                        Reset Delay                     True
rw_timeout      30                       READ/WRITE time out value       True
scsi_id         0xd22000                 SCSI ID                         False
start_timeout   60                       START UNIT time out value       True
timeout_policy  fail_path                Timeout Policy                  True
ww_name         0x50001442607ef401       FC World Wide Name              False
#
```

2. To change the queue depth value, run the **`chdev`** command for each disk.

   `chdev -l hdisk# -a queue_depth=<value>`

   For example, to set the queue depth value to 20 on the disk, use the following command:

   `chdev -l hdisk# -a queue_depth=20`

```
# chdev -l hdisk2 -a queue_depth=20
hdisk2 changed
#
```

3. To verify that the new queue depth value has been set, run the **`lsattr`** command as mentioned below.

```
lsattr -l hdisk# -E
```

```
# lsattr -l hdisk2 -E
PCM             PCM/friend/MINVISTA_DISK Path Control Module             True
PR_key_value    none                     Persistant Reserve Key Value    True
algorithm       round_robin              Algorithm                       True
clr_q           yes                      Device CLEARS its Queue on error True
dist_err_pcnt   0                        Distributed Error Percentage    True
dist_tw_width   50                       Distributed Error Sample Time   True
hcheck_cmd      test_unit_rdy            Health Check Command            True
hcheck_interval 60                       Health Check Interval           True
hcheck_mode     nonactive                Health Check Mode               True
location                                 Location Label                  True
lun_id          0x0                      Logical Unit Number ID          False
lun_reset_spt   yes                      FC Forced Open LUN              True
max_coalesce    0x100000                 Maximum Coalesce Size           True
max_retries     5                        Maximum Number of Retries       True
max_transfer    0x100000                 Maximum TRANSFER Size           True
node_name       0x5000144046e07ef4       FC Node Name                    False
pvid            none                     Physical volume identifier      False
q_err           no                       Use QERR bit                    True
q_type          simple                   Queue TYPE                      True
queue_depth     20                       Queue DEPTH                     True
reassign_to     120                      REASSIGN time out value         True
reserve_policy  no_reserve               Reserve Policy                  True
reset_delay     0                        Reset Delay                     True
rw_timeout      30                       READ/WRITE time out value       True
scsi_id         0xd22000                 SCSI ID                         False
start_timeout   60                       START UNIT time out value       True
timeout_policy  fail_path                Timeout Policy                  True
ww_name         0x50001442607ef401       FC World Wide Name              False
```

Setting num_cmd_elems

To set the `num_cmd_elems` on the Emulex Fibre Channel HBA, complete the following steps.

1. To check the existing `num_cmd_elems`, run the **`lsattr`** command for each HBA.

   ```
   lsattr -EHl fcs# -a num_cmd_elems
   ```

   Note: Replace `fcs#` with the HBA number (`fcs0`, `fcs1`, etc.).

```
# lsattr -EHl fcs0 -a num_cmd_elems
attribute      value description                                       user_settable

num_cmd_elems 1024  Maximum number of COMMANDS to queue to the adapter True
```

2. Run the following **`chdev`** command for each HBA in the AIX host.

   ```
   chdev -l fcs# -a num_cmd_elems=<value> -P
   ```

For example, to set the num_cmd_elems to 64 on the HBA use the following command.

chdev -l fcs# -a num_cmd_elems=64 -P

```
# chdev -l fcs0 -a num_cmd_elems=64 -P
fcs0 changed
```

3. Reboot the AIX host to apply the num_cmd_elems settings.
4. Once the host boots up, verify that the new num_cmd_elements value has been set using the following **lsattr** command.

   lsattr -EHl fcs# -a num_cmd_elems

```
# lsattr -EHl fcs0 -a num_cmd_elems
attribute      value description                                       user_settable

num_cmd_elems 64    Maximum number of COMMANDS to queue to the adapter True
```

Setting max_xfer_size and lg_term dma attributes

To set the max_xfer_size and lg_term_dma attributes on the Emulex Fibre Channel HBA, complete the following steps.

1. To view the value set for max_xfer_size, issue the following command.

   lsattr -EHl fcs# -a num_cmd_elems

   Note: Replace fcs# with the HBA number (fcs0, fcs1, etc.).

```
# lsattr -EHl fcs0 -a max_xfer_size
attribute      value      description           user_settable

max_xfer_size 0x1000000 Maximum Transfer Size True
```

2. To change the value for max_xfer_size, run the **chdev** command as shown below.

   chdev -l fcs# -a max_xfer_size=<value> -P

   For example, to set the max_xfer_size to 0x1000000 on the HBA use the following command.

   chdev -l fcs# -a max_xfer_size=0x1000000 -P

```
# chdev -l fcs0 -a max_xfer_size=0x1000000 -P
fcs0 changed
```

3. To view the value set for lg_term_dma, type the following command.

   lsattr -EHl fcs# -a lg_term_dma

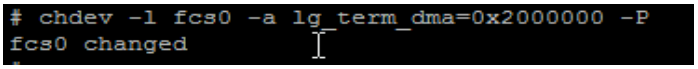
```
# lsattr -EHl fcs0 -a lg_term_dma
attribute   value     description   user_settable

lg_term_dma 0x2000000 Long term DMA True
```

4. To change the value for lg_term_dma, run the **chdev** command as shown below.

chdev -l fcs# -a lg_term_dma=<value> -P

For example, to set the lg_term_dma to 0x2000000 on the HBA use the following command.

chdev -l fcs# -a lg_term_dma=0x2000000 -P



5. Reboot the AIX host to apply the settings.
6. Once the host boots up, verify that the new `max_xfer_size` and `lg_term_dma` values have been set using the following **`lsattr`** command.

   lsattr -EHl fcs# -a max_xfer_size

Note: If you still want to improve IO performance, you can increase the value of `lg_term_dma` attribute again. However, If you have a dual-port Fibre Channel adapter, the maximum value of the `lg_term_dma` attribute is divided between the two adapter ports. Therefore, never increase the value of `lg_term_dma` attribute to the maximum value for a dual-port Fibre Channel adapter because this value causes the configuration of the second adapter to fail.

Provisioning and exporting storage

This section provides information for the following:

VPLEX with GeoSynchrony v4.x

To begin using VPLEX, you must provision and export storage so that hosts and applications can use the storage. Storage provisioning and exporting refers to the following tasks required to take a storage volume from a storage array and make it visible to a host:

1. Discover available storage.
2. Claim and name storage volumes.
3. Create extents from the storage volumes.
4. Create devices from the extents.
5. Create virtual volumes on the devices.
6. Create storage views to allow hosts to view specific virtual volumes.
7. Register initiators with VPLEX.
8. Add initiators (hosts), virtual volumes, and VPLEX ports to the storage view.

You can provision storage using the GUI or the CLI. Refer to the VPLEX Management Console Help or the *EMC VPLEX CLI Guide,* located at Dell EMC Online Support, for more information.

Figure 18 on page 363 illustrates the provisioning and exporting process.

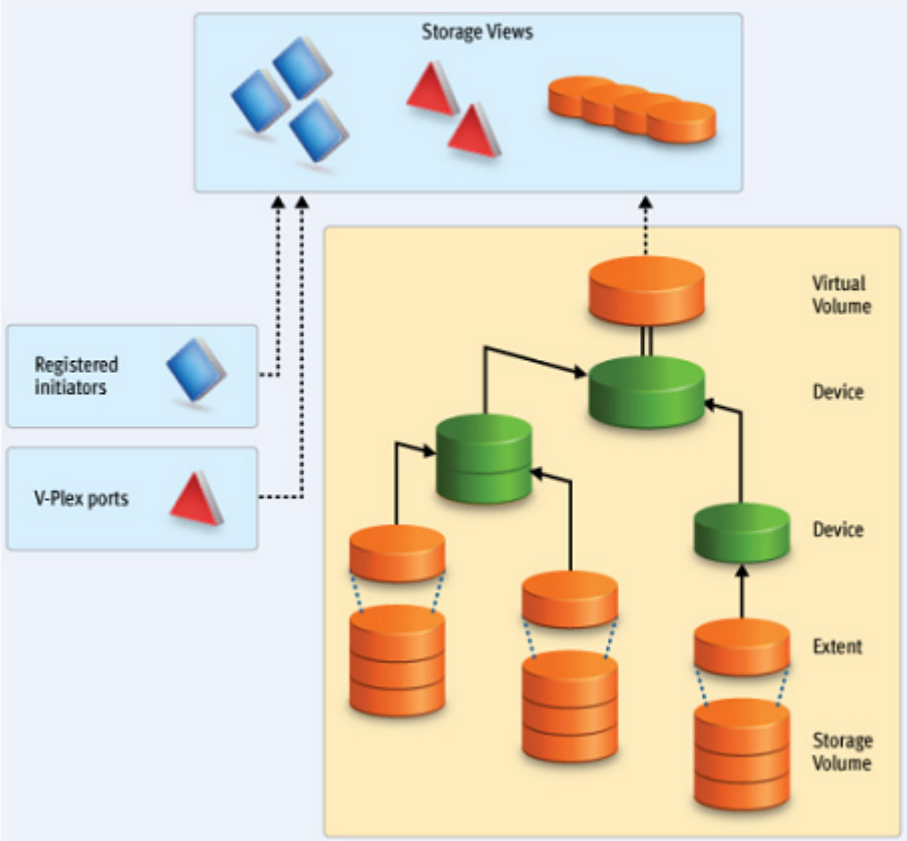


Figure 18 VPLEX provisioning and exporting storage process

VPLEX with GeoSynchrony v5.x

VPLEX allows easy storage provisioning among heterogeneous storage arrays. After a storage array LUN volume is encapsulated within VPLEX, all of its block-level storage is available in a global directory and coherent cache. Any front-end device that is zoned properly can access the storage blocks.

Two methods available for provisioning: EZ provisioning and Advanced provisioning. For more information, refer to the *EMC VPLEX with GeoSynchrony Product Guide*, located at Dell EMC Online Support.

Storage volumes

A storage volume is a LUN exported from an array. When an array is discovered, the storage volumes view shows all exported LUNs on that array. You must claim, and optionally name, these storage volumes before you can use them in a VPLEX cluster. Once claimed, you can divide a storage volume into multiple extents (up to 128), or you can create a single full size extent using the entire capacity of the storage volume.

Note: To claim storage volumes, the GUI supports only the Claim Storage wizard, which assigns a meaningful name to the storage volume. Meaningful names help you associate a storage volume with a specific storage array and LUN on that array, and helps during troubleshooting and performance analysis.

This section contains the following information:

Claiming and naming storage volumes

You must claim storage volumes before you can use them in the cluster (with the exception of the metadata volume, which is created from an unclaimed storage volume). Only after claiming a storage volume, can you use it to create extents, devices, and then virtual volumes.

Extents

An extent is a slice (range of blocks) of a storage volume. You can create a full size extent using the entire capacity of the storage volume, or you can carve the storage volume up into several contiguous slices. Extents are used to create devices, and then virtual volumes.

IMPORTANT

When creating multiple extents on a storage volume, care must be taken to ensure that the physical storage corresponding to the storage volume can sufficiently handle the IO load.

By default, when you create an extent, VPLEX uses the entire storage capacity of the selected storage volume. However, you can specify less than the entire capacity of the storage volume.

You can create a maximum of 128 extents per storage volume.

Note: If encapsulating a storage volume, create a single full size extent on the claimed storage volume to ensure that hosts can access all the data on the storage volume.

Devices

Devices combine extents or other devices into one large device with specific RAID techniques, such as mirroring or striping. Devices can only be created from extents or other devices. A device's storage capacity is not available until you create a virtual volume on the device and export that virtual volume to a host.

You can create only one virtual volume per device. There are two types of devices:

- *Simple device — A simple device is configured using one component, which is an extent.*
- *Complex device — A complex device has more than one component, combined using a specific RAID type. The components can be extents or other devices (both simple and complex).*

Distributed devices

Distributed devices are configured using storage from both clusters and therefore are only used in multi-cluster plexes. A distributed device's components must be other devices and those devices must be created from storage in different clusters in the plex.

Rule sets

Rule sets are predefined rules that determine how a cluster behaves when it loses communication with the other cluster, for example, during an inter-cluster link failure or cluster failure. In these situations, until communication is restored, most I/O workloads require specific sets of virtual volumes to resume on one cluster and remain suspended on the other cluster.

VPLEX provides a Management Console on the management server in each cluster. You can create distributed devices using the GUI or CLI on either management server. The default rule set used by the GUI makes the cluster used to create the distributed device detach during communication problems, allowing I/O to resume at the cluster. For more information, on creating and applying rule sets, refer to the *EMC VPLEX CLI Guide*, available at Dell EMC Online Support.

There are cases in which all I/O must be suspended resulting in a data unavailability. VPLEX with GeoSynchrony 5.0 is introduces the new functionality of VPLEX Witness. When a VPLEX Metro or a VPLEX Geo configuration is augmented by VPLEX Witness, the resulting configuration provides the following features:

- *High availability for applications in a VPLEX Metro configuration (no single points of storage failure)*
- *Fully automatic failure handling in a VPLEX Metro configuration*
- *Significantly improved failure handling in a VPLEX Geo configuration*
- *Better resource utilization*

For information on VPLEX Witness, refer the *EMC VPLEX with GeoSynchrony Product Guide*, located at Dell EMC Online Support.

Virtual volumes

Virtual volumes are created on devices or distributed devices and presented to a host via a storage view. Virtual volumes can be created only on top-level devices and always use the full capacity of the device.

System volumes

VPLEX stores configuration and metadata on system volumes created from storage devices. There are two types of system volumes. Each is briefly discussed in this section:

Metadata volumes

VPLEX maintains its configuration state, referred to as metadata, on storage volumes provided by storage arrays. Each VPLEX cluster maintains its own metadata, which describes the local configuration information for this cluster as well as any distributed configuration information shared between clusters.

For more information about metadata volumes for VPLEX with GeoSynchrony v4.x, refer to the *EMC VPLEX CLI Guide*, available at Dell EMC Online Support.

For more information about metadata volumes for VPLEX with GeoSynchrony v5.x, refer to the *EMC VPLEX with GeoSynchrony Product Guide*, located at Dell EMC Online Support.

Logging volumes

Logging volumes are created during initial system setup and are required in each cluster to keep track of any blocks written during a loss of connectivity between clusters. After an inter-cluster link is restored, the logging volume is used to synchronize distributed devices by sending only changed blocks over the inter-cluster link.

For more information about logging volumes for VPLEX with GeoSynchrony v4.x, refer to the *EMC VPLEX CLI Guide*, available at Dell EMC Online Support.

For more information about logging volumes for VPLEX with GeoSynchrony v5.x, refer to the *EMC VPLEX with GeoSynchrony Product Guide*, located at Dell EMC Online Support.

Required storage system setup

Symmetrix, VNX series, or CLARiiON product documentation and installation procedures for connecting a Symmetrix, VNX series, or CLARiiON storage system to a VPLEX instance are available at Dell EMC Online Support.

Required Symmetrix FA bit settings

For Symmetrix-to-VPLEX connections, configure the Symmetrix Fibre Channel directors (FAs) as shown in Table 16.

Note: Dell EMC recommends that you download the latest information before installing any server.

Table 16 Required Symmetrix FA bit settings

Set	Do not set	Optional
SPC-2 Compliance (SPC2) SCSI-3 Compliance (SC3) Enable Point-to-Point (PP) Unique Worldwide Name (UWN) Common Serial Number (C)	Disable Queue Reset on Unit Attention (D) AS/400 Ports Only (AS4) Avoid Reset Broadcast (ARB) Environment Reports to Host (E) Soft Reset (S) Open VMS (OVMS) Return Busy (B) Enable Sunapee (SCL) Sequent Bit (SEQ) Non Participant (N) OS-2007 (OS compliance)	Linkspeed Enable Auto-Negotiation (EAN) VCM/ACLX [1]

1. Must be set if VPLEX is sharing Symmetrix directors with hosts that require conflicting bit settings. For any other configuration, the VCM/ACLX bit can be either set or not set.

Note: When setting up a VPLEX-attach version 4.x or earlier with a VNX series or CLARiiON system, the initiator type must be set to CLARiiON Open and the Failover Mode set to 1. ALUA is not supported.

When setting up a VPLEX-attach version 5.0 or later with a VNX series or CLARiiON system, the initiator type can be set to CLARiiON Open and the Failover Mode set to 1 or Failover Mode 4 since ALUA is supported.

If you are using the LUN masking, you will need to set the VCM/ACLX flag. If sharing array directors with hosts which require conflicting flag settings, VCM/ACLX must be used.

Note: The FA bit settings listed in Table 16 are for connectivity of VPLEX to EMC Symmetrix arrays only. For Host to EMC Symmetrix FA bit settings, refer to the Dell EMC Simple Support Matrices.

Supported storage arrays

The *EMC VPLEX Simple Support Matrix* lists the storage arrays that have been qualified for use with VPLEX.

Refer to the *VPLEX Procedure Generator,* available at Dell EMC Online Support, to verify supported storage arrays.

VPLEX automatically discovers storage arrays that are connected to the back-end ports. All arrays connected to each director in the cluster are listed in the storage array view.

Initiator settings on back-end arrays

Refer to the *VPLEX Procedure Generator,* available at Dell EMC Online Support, to verify the initiator settings for storage arrays when configuring the arrays for use with VPLEX.

Host connectivity

For the most up-to-date information on qualified switches, hosts, host bus adapters, and software, refer to the always consult the Dell EMC Simple Support Matrices, available through Dell E-Lab Navigator, or contact your Dell EMC Customer Representative.

The latest approved HBA drivers and software are available for download at the following websites:

- *http://www.emulex.com*
- *http:/www.qlogic.com*
- *http://www.brocade.com*

The EMC HBA installation and configurations guides are available at the EMC-specific download pages of these websites.

Note: Direct connect from a host bus adapter to a VPLEX engine is not supported.

Exporting virtual volumes to hosts

A virtual volume can be added to more than one storage view. All hosts included in the storage view will be able to access the virtual volume.

The virtual volumes created on a device or distributed device are not visible to hosts (or initiators) until you add them to a storage view. For failover purposes, two or more front-end VPLEX ports can be grouped together to export the same volumes.

A volume is exported to an initiator as a LUN on one or more front-end port WWNs. Typically, initiators are grouped into initiator groups; all initiators in such a group share the same view on the exported storage (they can see the same volumes by the same LUN numbers on the same WWNs).

An initiator must be registered with VPLEX to see any exported storage. The initiator must also be able to communicate with the front-end ports over a Fibre Channel switch fabric. Direct connect is not supported. Registering an initiator attaches a meaningful name to the WWN, typically the server's DNS name. This allows you to audit the storage view settings to determine which virtual volumes a specific server can access.

Exporting virtual volumes consists of the following tasks:

1. Creating a storage view, as shown in Figure 19.

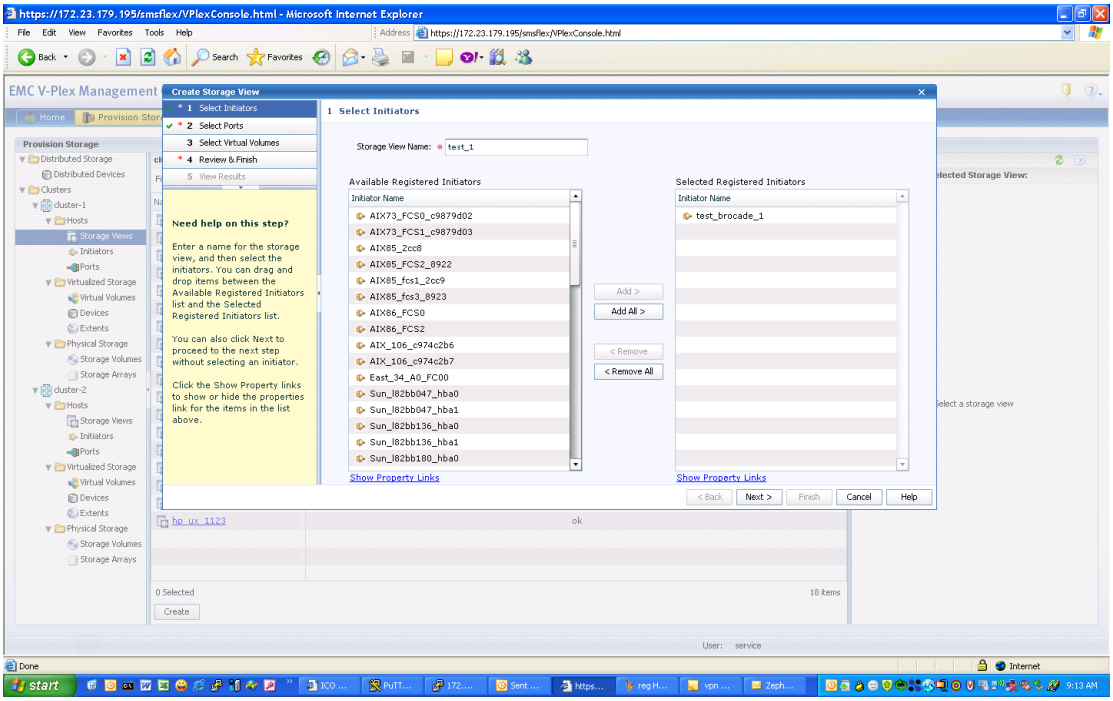


Figure 19 Create storage view

2. Registering initiators, as shown in Figure 20.

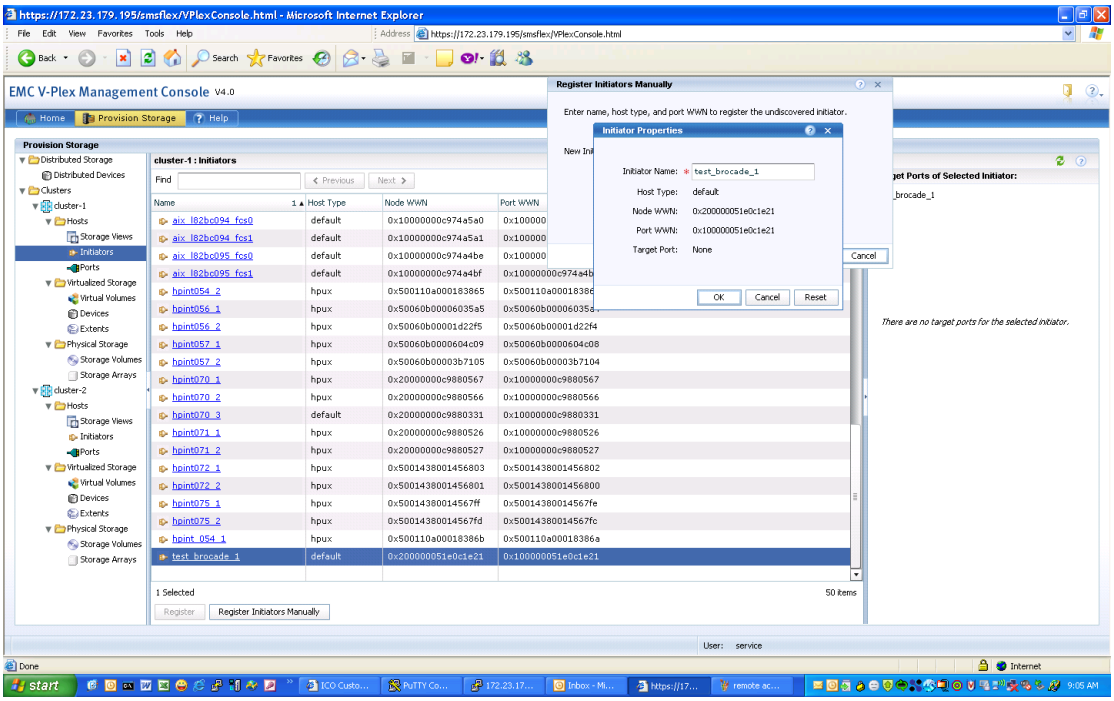


Figure 20 Register initiators

Note: When initiators are registered, you can set their type as indicated in Table 17 on page 375.

3. Adding ports to the storage view, as shown in Figure 21.

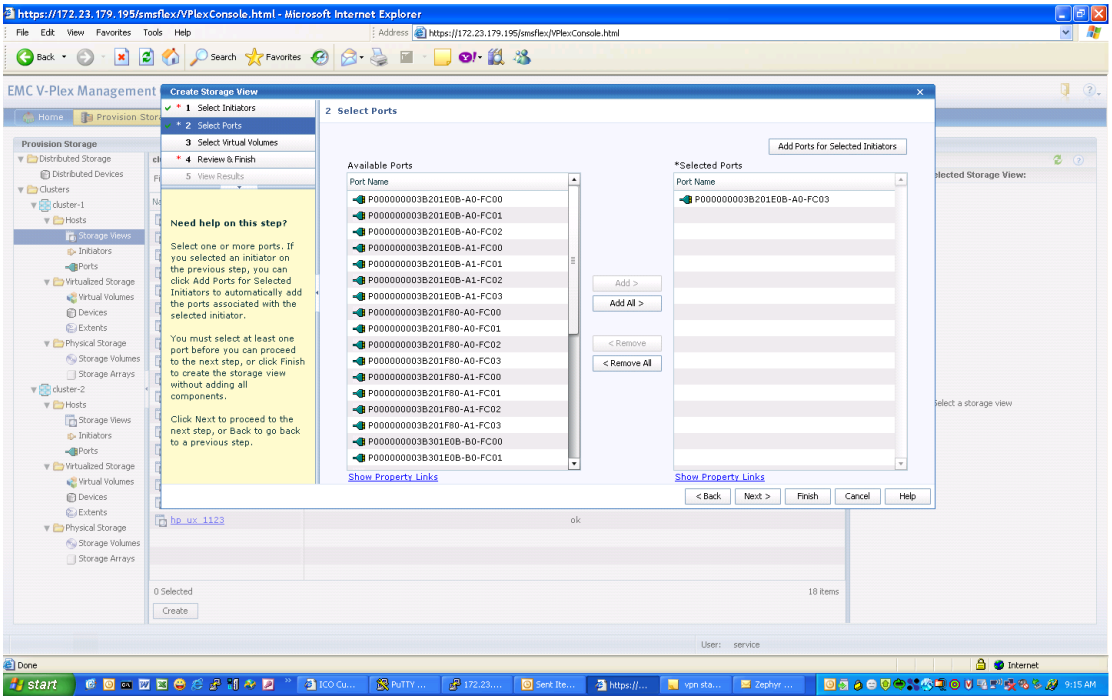


Figure 21 Add ports to storage view

4. Adding virtual volumes to the storage view, as shown in Figure 22.

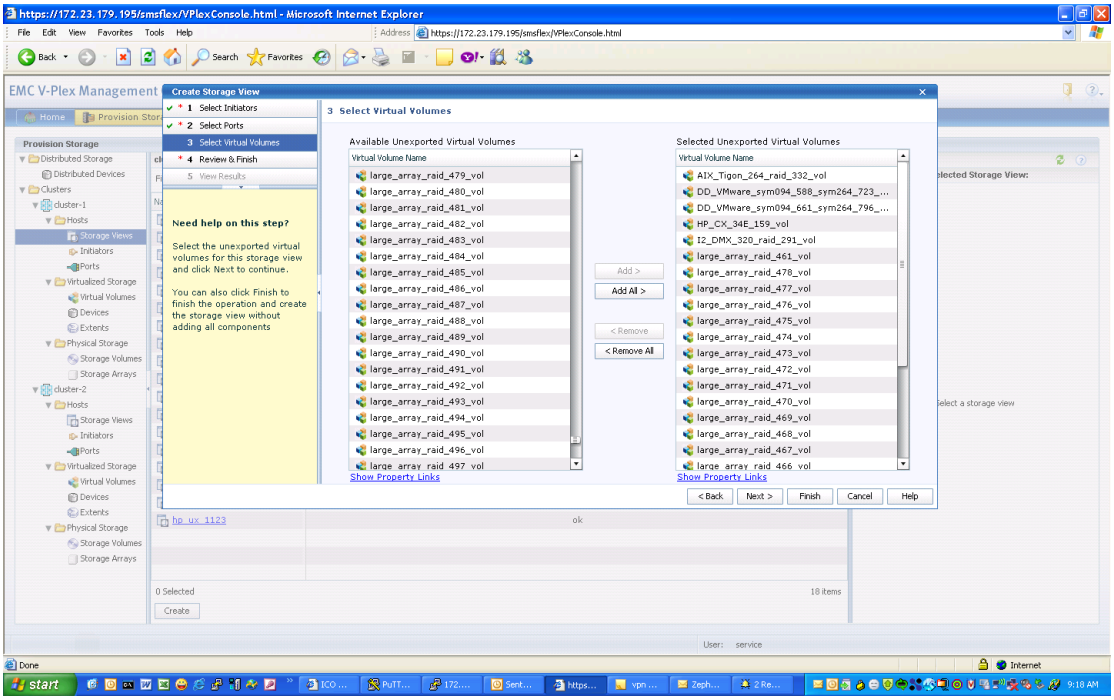


Figure 22 Add virtual volumes to storage view

Front-end paths

This section contains the following information:

Viewing the World Wide Name for an HBA port

This section describes how to view the WWNs of the Fibre Channel HBAs in an IBM AIX host. The following example shows two HBAs are found using the **lsdev -Cc adapter** command. The **lscfg** command is then used to extract the WWNs.

```
>lsdev -Cc adapter|grep fcs
fcs0 Available 2V-08 FC Adapter
fcs1 Available 2K-08 FC Adapter
>lscfg -v1 fcs0|grep Network
Network Address.............10000000C93AB6EB
>lscfg -v1 fcs1|grep Network
Network Address.............10000000C93AB5E4
```

VPLEX ports

The virtual volumes created on a device are not visible to hosts (initiators) until you export them. Virtual volumes are exported to a host through front-end ports on the VPLEX directors and HBA ports on the host/server. For failover purposes, two or more front-end VPLEX ports can be used to export the same volumes. Typically, to provide maximum redundancy, a storage view will have two VPLEX ports assigned to it, preferably from two different VPLEX directors. When volumes are added to a view, they are exported on all VPLEX ports in the view, using the same LUN numbers.

Initiators

For an initiator to see the virtual volumes in a storage view, it must be registered and included in the storage view's registered initiator list. The initiator must also be able to communicate with the front-end ports over Fibre Channel connections through a fabric.

A volume is exported to an initiator as a LUN on one or more front-end port WWNs. Typically, initiators are grouped so that all initiators in a group share the same view of the exported storage (they can see the same volumes by the same LUN numbers on the same WWN host types).

Ensure that you specify the correct host type in the **Host Type** column as this attribute cannot be changed in the **Initiator Properties** dialog box once the registration is complete. To change the host type after registration, you must unregister the initiator and then register it again using the correct host type.

VPLEX supports the host types listed in Table 17. When initiators are registered, you can set their type, also indicated in Table 17.

Table 17 Supported hosts and initiator types

Host	Initiator type
Windows, MSCS, Linux	default
AIX	Aix
HP-UX	Hp-ux
Solaris, VCS	Sun-vcs

Configuring IBM AIX to recognize VPLEX volumes

You must configure an IBM AIX host to recognize VPLEX Virtual Volumes. To do this, install the VPLEX ODM filesets by completing the following steps:

Note: VPLEX will present its volumes to the host with the device type "Invista" as the Invista array. Therefore, in this section, to be consistent with what you will see in the example outputs, "Invista devices" or "Invista" will be used when referring to devices that VPLEX present to the hosts.

1. Log in to the host as root.
2. Remove all devices from the host configuration.
3. Obtain the ODM package from the EMC-FTP server:

   ```
   ftp ftp.emc.com
   cd/pub/elab/aix/ODM_DEFINITIONS
   bin
   get EMC.AIX.5.3.0.x.tar.Z
   ```

   Note: Use the most recent fileset available.

   The ODM readme file lists the supported versions of AIX.

4. In the /tmp directory, uncompress the tar package:

   ```
   uncompress EMC.AIX.5.3.0.X.tar.Z
   ```

5. Extract the tar package:

   ```
   tar -xvf EMC.AIX.5.3.0.X.tar
   ```

6. Perform the remaining steps using either the command line or SMIT:

   To use the command line:

   a. Type the following commands:

   Note: The period between -d and EMC.Invista specifies the current directory.

   If using PowerPath:

   ```
   installp -ac -gX -d . EMC.Invista.aix.rte
   installp -ac -gX -d . EMC.Invista.fcp.rte
   ```

   If using MPIO:

   ```
   installp -ac -gX -d . EMC.Invista.aix.rte
   installp -ac -gX -d . EMC.Invista.fcp.MPIO.rte
   ```

   b. Verify that the installation summary reports SUCCESS.

   If the installation is successful, proceed to Step i on page 377.

   To use SMIT:

   a. Type the following command line:

   ```
   smitty installp
   ```

b. Select **Install and Update form Latest Available Software**.

c. Select the input device to be the current working directory:

```
directory
./
```

d. Select **F4=List** to list the available software.

e. Select **F7=Select** to select EMC Invista AIX Support Software and EMC Invista Fibre Channel Support Software.

f. Select **Enter**.

g. Accept the default options and press **Enter** to initiate the installation.

h. Scroll to the bottom of the installation summary screen to verify that the SUCCESS message is displayed.

i. Exit SMIT.

7. Choose one of the following:

- For PowerPath hosts:

  Use the emc_cfgmgr -v script to automatically configure all paths and then update the PowerPath configuration:

```
>/usr/lpp/EMC/INVISTA/bin/emc_cfgmgr -v
>powermt config
```

- For DMP hosts:

  Use the cfgmgr -v to automatically configure all paths and then vxdisk scan disks to configure DMP:

```
>cfgmgr -v
>vxdisk scandisks
```

8. Choose one of the following:

- For PowerPath hosts:

  Verify that all hdisks have been configured properly down each path:

```
>lsdev -Cc disk
...
hdisk2 Available 2V-08-01 EMC INVISTA FCP Disk
hdisk3 Available 2K-08-01 EMC INVISTA FCP Disk
...
hdiskpower0 Available 2V-08-01 PowerPath Device
hdiskpower1 Available 2V-08-01 PowerPath Device
```

- For DMP hosts:

  Verify that all hdisks have been configured down each path and that DMP is recognizing the VPLEX array:

```
>lsdev -Cc disk
...
hdisk2 Available 2V-08-01 EMC INVISTA FCP Disk
hdisk3 Available 2K-08-01 EMC INVISTA FCP Disk
...

>vxdmpadm getsubpaths dmpnodename=hdisk2
```

```
NAME          STATE[A]   PATH-TYPE[M] CTLR-NAME  ENCLR-TYPE   ENCLR-NAME    ATTRS
================================================================================
hdisk16       ENABLED(A)    -          fscsi1     EMC-VPLEX    emc-vplex0       -
hdisk2        ENABLED(A)    -          fscsi0     EMC-VPLEX    emc-vplex0       -
```

>**vxdmpadm listenclosure**

```
ENCLR_NAME        ENCLR_TYPE     ENCLR_SNO      STATUS       ARRAY_TYPE     LUN_COUNT
================================================================================
emc-vplex0        EMC-VPLEX      FNM00103700051       CONNECTED    VPLEX-A/A  3
```

PART 5

Data Protection

Part 6 includes:

CHAPTER 13

Data Domain Support

This chapter is a reference for Dell EMC Data Domain support in the IBM AIX environment.

Overview

Dell EMC Data Domain is a protection storage platform, which consolidates backup, archive, and disaster recovery all on a single system. By leveraging variable-length deduplication, Data Domain systems are able to reduce disk storage by 10-30x or greater, making disk a cost-effective alternative to tape. For more details, refer to Dell EMC Data Domain documentation, available at https://support.emc.com.

For configuration and administration operations, refer to Dell EMC Simple Support Matrix, available at https://elabnavigator.emc.com/eln/elnhome.

Prerequisites

Before configuring VPLEX in the IBM AIX environment, perform the following on each host:

Host connectivity

Refer to the Dell EMC Support Matrix or contact your Dell EMC representative for the latest information on qualified hosts, host bus adapters, and connectivity equipment.

- Support AIX version: AIX 6.1, AIX7.1 and AIX 7.2.

ODM version requirement

Data Domain DFC (DDBoost over Fiber Channel) MPIO disk can be supported on AIX system:

- AIX 6.1 requires EMC ODM 5.3.1.2 or later.
- AIX 7.1 and AIX 7.2 require EMC ODM 6.0.0.6 or later.

DFC MPIO Connection Configuration

Install the Data Domain ODM filesets to configure the AIX host to use Data Domain DFC disks. Refer to following steps for Data Domain DFC disk configuration:

Installing EMC ODM kit

1. Enter the following command to remove all the Data Domain disks, if they're recognized as **Other FC SCSI Disk Drive** by system:

   ```
   lsdev -Cc disk | grep "Other FC SCSI Disk Drive" | awk
   {'print $1'} | xargs -n1 rmdev -dl
   ```

1. Download the EMC AIX ODM Package Version 5.3.x.x or 6.0.x.x from the EMC ftp server ftp.emc.com.

   a. To download the EMC-supplied AIX ODM Package, type the following command:

      ```
      ftp ftp.emc.com
      ```

   a. Log in with a username of anonymous and your email address as a password.

   b. Type the following command:

      ```
      cd /pub/elab/aix/ODM_DEFINITIONS
      ```

   c. Type the following and set the file transfer mode to binary:

      ```
      binary
      ```

   d. Type the following command:

      ```
      get EMC.AIX.5.3.x.x.tar.Z
      ```

   e. Type the following command:

      ```
      bye
      ```

1. Save the EMC AIX ODM Package Version 5.3.x.x or 6.0.x.x into the /usr/sys/inst.images directory for installation.

2. To uncompress and untar the EMC AIX ODM package type the following:

   ```
   cd /usr/sys/inst.images
   uncompress EMC.AIX.5.3.x.x.tar.Z
   tar -xvf EMC.AIX.5.3.x.x.tar
   ```

3. To Install the desired EMC Data Domain ODM filesets, type the following:

   ```
   rm .toc
   inutoc .
   ```

4. To Install the desired EMC Data Domain ODM filesets using DFC MPIO disks, type the following:

   ```
   installp -ac -gX -d . EMC.DataDomain.aix.rte
   installp -ac -gX -d . EMC.DataDomain.fcp.MPIO.rte
   ```

Note: You can also use **smitty** to install the EMC Data Domain ODM filesets.

Installing EMC ODM kit

5. To scan the disks, type the following AIX command:

   cfgmgr

6. Verify that all disks have been configured properly.
7. To view the recognized Data Domain DFC disks, type the following command:

   lsdev -Cc disk

```
hdisk2   Available 00-T1-01 EMC DataDomain DFC MPIO Disk
hdisk3   Available 00-T1-01 EMC DataDomain DFC MPIO Disk
```

PART 6

Appendix

Part 7 includes:

APPENDIX A

Migrating Considerations for AIX

Refer to this appendix when migrating AIX from one Dell EMC array to another. This appendix contains information that may be helpful when encountering problems with AIX, attached to the new array.

Note: Any general references to Symmetrix or Symmetrix array include the VMAX3 Family, VMAX, and DMX.

AIX checklist

- Refer to "Useful AIX functions and utilities" on page 25 for specific AIX commands necessary to execute an EMC array migration process. For example, the **`cfgmgr`** command is required to discover the new array.
- Chapter 1, "General AIX Information," contains specific steps necessary for the configuration of a new Dell EMC array.
- Prior to the migration, capture an EMCGrab, IBM Snap, and VRTSexplorer (if running VxVM) output for each AIX server for your records.

VMAX or Symmetrix checklist

- Ensure the FA and/or SE director bit settings are consistent between each VMAX or Symmetrix.

  In some cases, new Symmetrix arrays and new Symmetrix Microcode require different director bit settings. Refer to the Dell EMC Support Matrices for the proper director bit settings, for AIX initiators.
- On each switch, ensure each zone has the correct WWN for all appropriate HBAs and Symmetrix ports.
- When using Symantec DMP for failover, ensure the “C” bit is enabled on all appropriate FA ports.

VNX series or CLARiiON checklist

- Using Unisphere/Navisphere, verify that the 'Initiator Type', 'failovermode' and 'Arraycommpath' are all properly set for each AIX initiator.
- Verify each AIX initiator is "Registered" and "Logged In", using the **Connectivity Status** tool in Unisphere/Navisphere.
- On each switch, ensure each zone has the correct WWN for all appropriate HBAs and VNX series or CLARiiON ports.
- When using Symantec DMP for failover, verify the 'failovermode' is set to **1** for each AIX initiator.