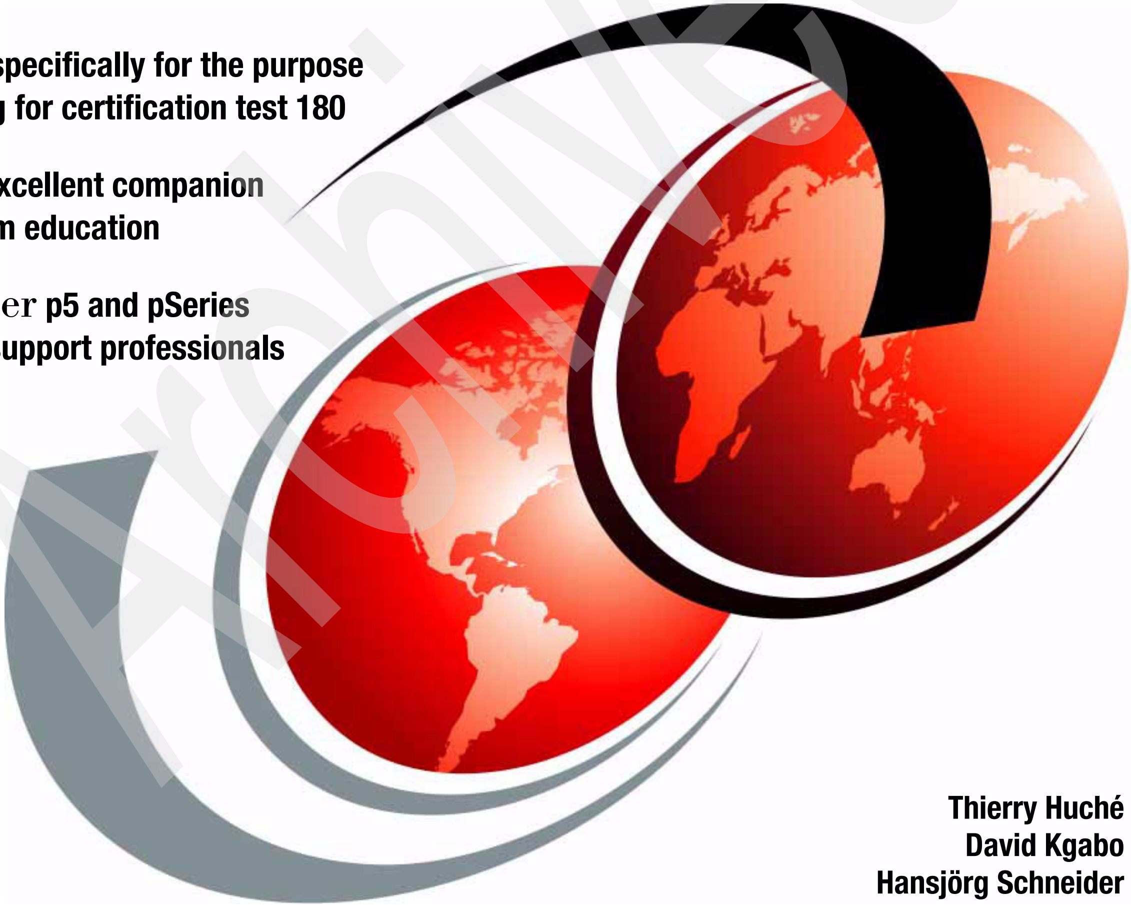
IBM

IBM Certified ™
Specialist

IBM @server Certification Study Guide @server p5 and pSeries Enterprise Technical Support AIX 5L V5.3



Developed specifically for the purpose of preparing for certification test 180

Makes an excellent companion to classroom education

For @server p5 and pSeries enterprise support professionals

Thierry Huché
David Kgabo
Hansjörg Schneider

Redbooks

ibm.com/redbooks

International Technical Support Organization

IBM @server Certification Study Guide: @server p5 and pSeries Enterprise Technical Support AIX 5L V5.3

December 2005

SG24-7197-01

Note: Before using this information and the product it supports, read the information in "Notices" on page ix.

Second Edition (December 2005)

This edition applies to IBM eserver p5 and pSeries enterprise servers for use with the IBM AIX 5L Version 5.3 operating system (program number 5765-G03).

Contents

Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:
IBM Director of Licensing, IBM Corporation, North Castle Drive Armonk, NY 10504-1785 U.S.A.

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurement may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:
This information contains sample application programs in source language, which illustrates programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in

any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. You may copy, modify, and distribute these sample programs in any form without payment to IBM for the purposes of developing, using, marketing, or distributing application programs conforming to IBM's application programming interfaces.

Trademarks

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

AIX 5L™
AIX®
DB2®
Electronic Service Agent™
Enterprise Storage Server®
ESCON®
@server®
@server®
eServer™
GDPS®
Geographically Dispersed
Parallel Sysplex™
HACMP™
i5/OS®
ibm.com®
IBM®
iSeries™
LoadLeveler®
Micro-Partitioning™
OpenPower™
Parallel Sysplex®
POWER4™
POWER5™
POWER™
pSeries®
PTX®
Redbooks (logo) ™
Redbooks™
RS/6000®
System p5™
Tivoli®
TotalStorage®
Virtualization Engine™

The following terms are trademarks of other companies:

Java, JavaScript, and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.

Microsoft, Windows, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

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

Linux is a trademark of Linus Torvalds in the United States, other countries, or both.

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

Preface

The IBM® AIX® 5L™ and IBM System p5™ Certifications, offered through the Professional Certification Program from IBM, are designed to validate the skills required of technical professionals who work in the powerful and often complex IBM AIX 5L and IBM System p5 environments. A complete set of professional certifications are available. They include:

- IBM Certified Specialist - ⓔserver p5 and pSeries® Administration and Support for AIX 5L V5.3
- IBM ⓔserver Certified Specialist - AIX Basic Operations V5
- IBM ⓔserver Certified Specialist - p5 and pSeries Technical Sales Support
- IBM Certified Systems Expert - ⓔserver p5 and pSeries Enterprise Technical Support AIX 5L V5.3
- IBM ⓔserver Certified Systems Expert - pSeries Cluster 1600 PSSP V5
- IBM ⓔserver Certified Systems Expert - pSeries HACMP™ for AIX 5L
- IBM ⓔserver Certified Advanced Technical Expert - pSeries and AIX 5L

Each certification is developed by following a thorough and rigorous process to ensure that the exam is applicable to the job role and is a meaningful and appropriate assessment of skill. Subject matter experts who successfully perform the job participate throughout the entire development process. They bring a wealth of experience into the development process, making the exams much more meaningful than the typical test that only captures classroom knowledge and ensuring that the exams are relevant to the *real world*. Thanks to their effort, the test content is both useful and valid. The result of this certification is the value of the appropriate measurements of the skills required to perform the job role.

This IBM Redbook is designed as a study guide for professionals wanting to prepare for the certification exam to achieve IBM Certified Systems Expert - ⓔserver p5 and pSeries Enterprise Technical Support AIX 5L V5.3. This technical support certification validates a broad scope of configuration, installation, and planning skills. In addition, it covers administrative and diagnostic activities needed to support logical partitions and virtual resources.

This publication helps IBM ⓔserver® p5 and pSeries professionals seeking a comprehensive and task-oriented guide for developing the knowledge and skills required for the certification. It is designed to provide a combination of theory and practical experience needed for a general understanding of the subject matter.

This publication does not replace the practical experience you should have, but is an effective tool that, when combined with education activities and experience, should prove to be a very useful preparation guide for the exam. Due to the close association with the certification content, this publication might reflect older software and firmware levels of the IBM @server p5 systems and available features. If you are planning to take the @server p5 and pSeries Enterprise Technical Support AIX 5L V5.3 certification exam, this redbook is for you.

For additional information about certification and instructions about how to register for an exam, visit our Web site at:

`http://www.ibm.com/certify`

The team that wrote this redbook

This redbook was produced by a team of specialists from around the world working at the International Technical Support Organization, Austin Center.

Thierry Huché is an IT Specialist working at the Products and Solutions Support Center in Montpellier, France, as an IBM @server pSeries Benchmark Manager. He has worked at IBM France for 16 years, and he has more than 14 years of AIX System Administration, RS/6000®, and @server p5 and pSeries experience working in the pSeries Benchmark Center and AIX support center in Paris. He is an IBM Certified pSeries AIX System Administrator. His areas of expertise include benchmarking, performance tuning, networking, high availability, and problem determination. He coauthored the IBM Redbook *Communications Server for AIX Explored*, SG24-2591.

David Kgabo is a Advisory IT Support Specialist in South Africa. He has six years of experience in UNIX® and @server p5 and pSeries administration and support for AIX 5L. His areas of expertise include UNIX system support.

Hansjörg Schneider is an IBM System p5, @server p5, and pSeries Hardware Support Specialist at the IBM International Technical Support Central Region Front Office in Ehningen, Germany. He has 11 years of experience in the System p5, @server p5, and pSeries, and AIX 5L fields. He has worked at IBM Germany for 17 years. He supports clients and customer engineers to provide critical support related to IBM UNIX systems. His areas of expertise include hardware and SAN, and he is also an AIX Certified Specialist.

Thanks to the following people for their contributions to this project:

Grace Bauer
IBM Seattle

Gary Moehnke, Donald Heller
IBM Rochester

Jerry J. Petru
IBM Minneapolis

Become a published author

Join us for a two- to six-week residency program! Help write an IBM Redbook dealing with specific products or solutions, while getting hands-on experience with leading-edge technologies. You'll team with IBM technical professionals, Business Partners and/or clients.

Your efforts will help increase product acceptance and client satisfaction. As a bonus, you'll develop a network of contacts in IBM development labs, and increase your productivity and marketability.

Find out more about the residency program, browse the residency index, and apply online at:

ibm.com/redbooks/residencies.html

Comments welcome

Your comments are important to us!

We want our Redbooks™ to be as helpful as possible. Send us your comments about this or other Redbooks in one of the following ways:

- Use the online **Contact us** review redbook form found at:

  ibm.com/redbooks

- Send your comments in an email to:

  redbook@us.ibm.com

- Mail your comments to:

  IBM Corporation, International Technical Support Organization
  Dept. JN9B Building 905
  11501 Burnet Road
  Austin, Texas 78758-3493

1

Certification overview

This chapter provides an overview of the skill requirements needed to obtain an IBM Certified Systems Expert certification. The following sections are designed to provide a comprehensive review of specific topics that are essential for obtaining the certification IBM Certified Systems Expert - @server p5 and pSeries Enterprise Technical Support AIX 5L V5.3.

This level certifies an advanced level of pSeries and AIX 5L knowledge and understanding, both in breadth and depth. It verifies the ability to perform in-depth analysis, apply complex AIX 5L concepts, and provide resolution to critical problems, all in a variety of areas within AIX 5L including the hardware that supports it.

1.1 Certification requirements

To attain the IBM Certified Systems Expert - @server p5 and pSeries Enterprise Technical Support AIX 5L V5.3 certification, you must pass two tests:

- One test must be from the list in the following Required prerequisite section.
- The other test must be the @server p5 and pSeries Enterprise Technical Support AIX 5L V5.3 exam (Test 180).

1.1.1 Required prerequisite

Prior to attaining the IBM Certified Systems Expert - @server p5 and pSeries Enterprise Technical Support AIX 5L V5.3 certification, candidates must have the following certification:

- IBM Certified Specialist - @server p5 and pSeries Administration and Support for AIX 5L V5.3 (Test 222)

1.1.2 Recommended prerequisites

The following prerequisites are recommended:

- Minimum of six months experience in implementing or supporting enterprise systems.
- Both Test 180 and Test 222 are required for IBM Certified Systems Expert - @server p5 and pSeries Enterprise Technical Support AIX 5L V5.3.
- Either Test 197 or Test 232 can be combined with Test 222 or Test 180, respectively, but the V5.2 level will be indicated on this certification until the corresponding V5.3 tests are earned.

1.1.3 Information and registration for the certification exam

For the latest certification information, see the following Web site:

```
http://www.ibm.com/certify
```

1.1.4 Core requirements

Proficiency is established upon passing the following tests.

eServer p5 and pSeries Administration and Support for AIX 5L

Test 222 was developed for the AIX 5L V5.3 level certification.

eServer p5 and pSeries Enterprise Technical Support AIX 5L

Test 180 was developed for this certification.

Preparation for this exam is the subject of this publication. To pass the test, you need knowledge about the following topics.

Section I - System Planning and Design

To plan for the system:

- Determine detailed information about the applications (for example, operating system dependency and compatibility).
- Determine operating system requirements (AIX 5L V5.2, AIX 5L V5.3, Linux®, or IBM i5/OS®).
- Evaluate performance data (such as average and peak utilizations).
- Define sizing, capacity, performance, and growth requirements of the solution.
- Determine site and hardware requirements (such as power, cooling, space, cabling, and rack).
- Identify application availability requirements.
- Determine backup and recovery requirements.
- Determine LPAR profile requirements (such as dynamic LPAR, virtual I/O, Micro-Partitioning™, adapter, CPU, and memory).
- Determine virtual input/output (I/O) requirements.
- Determine load manager requirements (for example, Workload Manager and Partition Load Manager).
- Design the hardware and software solution and position appropriate alternatives.
- Validate the proposed solution.
- Determine provisioning requirements (for example, memory and CPU).
- Determine installation and system management requirements.
- Determine specific administrative parameters.

Section II - Installation and Configuration

To install and configure the system:

- Verify that components for installation match the order, and install plans have not changed.
- Assign boot devices for the machine or LPAR.
- Implement an Advanced System Management Interface installation.

- Install the OS using an appropriate method (for example, SAN, NIM, or alternate disk installation).
- Configure and validate DLPAR and remote commands, including Web-based System Manager and SSH.
- Verify network parameters and connectivity.
- Configure initial tuning parameters (such as SMT).
- Configure appropriate volume groups, logical volumes, and file systems.
- Determine when to select 32-bit instead of the 64-bit kernel.

Section III - Hardware Management

To manage the hardware:

- Prepare the server for hardware additions and removals.
- Select resource dependencies (such as the Hardware Management Console, or HMC, and server).
- Determine LPAR dependencies.
- Manage server firmware.
- Configure IBM Electronic Service Agent™.
- Maintain HMC software and BIOS versions.
- Plan and perform HMC backup and recovery.

Section IV - Ongoing Support

To perform ongoing support:

- Perform preventative hardware maintenance.
- Monitor and tune system performance.
- Manage system software updates.
- Isolate causes of failure and determine appropriate actions.
- Plan and perform system backup and recovery.
- Plan and perform disk management.

Section V - Virtualization

To install and optimize virtualization:

- Identify the advantages of Advanced POWER™ Virtualization features for pSeries enterprise servers (such as Partition Load Manager and Enterprise Workload Manager, virtual LAN, virtual SCSI, virtual CPU, and tools to reduce requirements for physical adapters).

- Implement Virtual I/O (VIO) Servers (such as, virtual Ethernet, Shared Ethernet Adapters, virtual SCSI, memory requirements, CPU requirements, physical adapter requirements, multiple VIO Servers).
- Install VIO Servers (for example, using HMC, using CD-ROM, or using NIM).
- Maintain VIO Servers (such as maintain VIO Server software, OEM drivers, and applications).
- Configure and maintain Micro-Partitioned LPARs (such as fractional CPU LPAR allocation, CPU pooling shared processor pool, virtual processors, capped, or weighted, and uncapped).
- Determine client dependencies to VIO Servers.

1.2 Certification education courses

Courses are offered to help you prepare for the certification tests. For a current list, visit the following Web site:

```
http://www.ibm.com/certify/tests/edu180.shtml
```

2

System planning and design

This chapter provides information about system planning and design for the following systems:

- IBM eserver p5 570
- IBM eserver p5 575
- IBM eserver p5 590 and p5-595

2.1 p5-570

The IBM @server p5 570 rack-mount server is designed for greater application flexibility, with innovative technology, to capitalize on the e-business revolution at the midrange level for server environments. With IBM POWER5™ microprocessor technology, the 9117 p5-570 is the first cost-effective, high-performance midrange UNIX server to include the next development of the IBM partitioning concept, Micro-Partitioning.

Dynamic logical partitioning is supported from the 2-way p5-570 to the 16-way p5-570 system, allowing up to 16 dedicated partitions. In addition, the optional Advanced POWER Virtualization hardware feature enables a technology called Micro-Partitioning technology. The p5-570 system has been designed to support up to 160 partitions on a 16-way system. The Micro-Partitioning technology is an advanced feature of the POWER5 processor that enables multiple partitions to share a physical processor. The extended POWER Hypervisor controls dispatching the physical processors to each of the partitions using Micro-Partitioning technology. In addition to Micro-Partitioning technology, the Advanced POWER Virtualization feature enables sharing of network and storage adapters to satisfy the I/O requests of partitions that do not have a dedicated physical I/O adapter.

In combination with the extraordinary POWER5 processor, the Micro-Partitioning technology is designed to increase system management efficiency and lower operating expenses through the multiple use of single physical resources that are installed in the p5-570 system.

Simultaneous multithreading, a standard feature of POWER5 technology, enables two threads to be executed at the same time on a single processor. Simultaneous multithreading is user-selectable with dedicated or processors from a shared pool for use by partitions using Micro-Partitioning technology.

The symmetric multiprocessor (SMP) p5-570 system features base 2-way, 4-way, 8-way, 12-way, and 16-way, 64-bit, copper-based, SOI-based POWER5 microprocessors running at 1.5 GHz, 1.65 GHz, and 1.9 GHz with 36 MB off-chip Level 3 cache configurations. The system is based on a concept of system building blocks. The p5-570 building blocks are facilitated with the use of processor interconnect and system SP Flex cables that enable as many as four 4-way p5-570 building blocks to be connected to achieve a true 16-way SMP combined system. Additional processor configurations are possible with the installation of IBM @server Capacity on Demand (CoD) features. Main memory starting at 2 GB can be expanded to 128 GB in a single drawer, based on the available DIMMs, for higher performance and exploitation of 64-bit addressing to meet the demands of enterprise computing, such as large database applications.

One p5-570 building block includes six hot-plug PCI-X[1] slots with Enhanced Error Handling (EEH) and an enhanced blind-swap mechanism, two Ultra320 SCSI controllers, one 10/100/1000 Mbps integrated dual-port Ethernet controller, two service processor communications ports, two USB 2.0 ports, two HMC ports, two remote RIO-2 ports, and two System Power Control Network (SPCN) ports.

The p5-570 includes two 3-pack front-accessible, hot-swap-capable disk bays. The six disk bays of one IBM eserver p5 570 building block can accommodate up to 880.8 GB of disk storage using the 146.8 GB Ultra320 SCSI disk drives. Two additional media bays are used to accept optional slimline media devices, such as DVD-ROM or DVD-RAM drives. The p5-570 also has I/O expansion capability using the RIO-2 bus, which enables attachment of the 7311 Model D10, 7311 Model D11, and 7311 Model D20 I/O drawers.

Additional reliability and availability features include redundant hot-plug cooling fans and redundant power supplies. Along with these hot-plug components, the p5-570 is designed to provide an extensive set of reliability, availability, and serviceability (RAS) features that include improved fault isolation, recovery from errors without stopping the system, avoidance of recurring failures, and predictive failure analysis.

2.1.1 System specifications

Table 2-1 lists the general system specifications of a single p5-570 drawer.

Table 2-1 p5-570 specifications

Description	Range
Operating temperature	5 to 35 degrees Celsius (41 to 95° Fahrenheit)
Relative humidity	8% to 80%
Maximum wet bulb	23 degrees C (73° F) (operating)
Noise level	6.5 bels (operating)
Operating voltage	200 to 240 V AC 50/60 Hz
Maximum power consumption	1,300 watts (maximum)
Maximum power source loading	1.37 kVA (maximum configuration)
Maximum thermal output	4,437 Btu[a]/hr (maximum configuration)

[a] British thermal unit (Btu)

[1] PCI stands for Peripheral Component Interconnect, and the X stands for extended performances.

2.1.2 Physical package

One p5-570 drawer is packaged in a 4U[2] rack-mounted enclosure, and it is available only in the rack-mounted form factor. Table 2-2 shows the major physical attributes found on the p5-570 building block.

Table 2-2 Physical packaging of the p5-570

Dimension	One p5-570 building block
Height	174.1 mm (6.85 in.)
Width	483 mm (19.0 in.)
Depth	790 mm (31.1 in.)
Weight	63.6 kg (140 lb)

Using the p5-570 building block, an installed system can be made of one to four building blocks. To help ensure the installation and serviceability in non-IBM, industry-standard racks, review the vendor's installation planning information for any product-specific installation requirements. The processor and SP Flex cables present an additional planning requirement.

Figure 2-1 on page 11 shows an p5-570 drawer.

[2] One Electronic Industries Association Unit (1U) is 44.45 mm (1.75 in.).

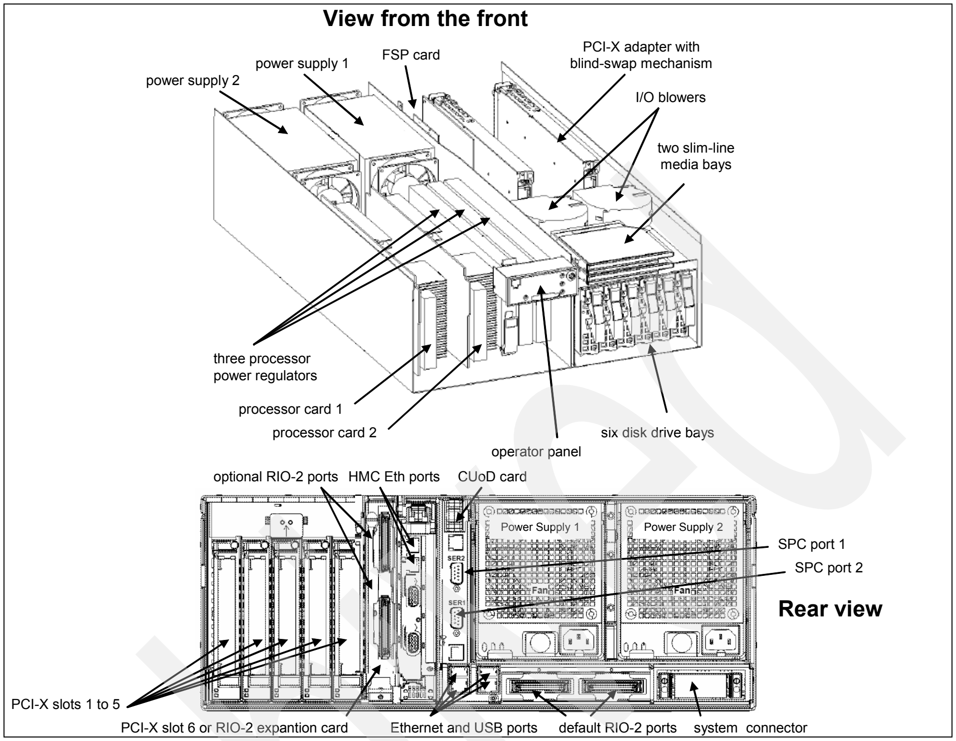


Figure 2-1 Views of the p5-570

2.1.3 Minimum and optional features

The p5-570 full configuration system is made of four p5-570 building blocks. It features:

- Up to eight processor books using the POWER5 chip, for a total of 16 processors
- From 2 GB to 512 GB of total system memory capacity using DDR1 DIMM technology, or from 2 GB to 64 GB total memory with DDR2 DIMM technology in a four-drawer system
- 24 SCSI disk drives for an internal storage capacity of 3.5 TB using 146.8 GB drives
- 24 PCI-X slots
- Eight slimline media bays for optional optical storage devices

The combined system (made up of more than one p5-570 building block) requires the proper processor interconnect cable and the system SP Flex cable.

The p5-570 building block includes the following native ports:

- Two 10/100/1000 Ethernet ports
- Two system ports
- Two USB 2.0 ports

  Optional external USB diskette drive 1.44 can be used.
- Two HMC ports
- Two remote I/O (RIO-2) ports
- Two SPCN ports

In addition, the p5-570 building block features two internal Ultra320 SCSI controllers, redundant hot-swap power supply and redundant hot-swap cooling fans, and redundant processor power regulators.

There is a CoD card as part of the hardware configuration. This card stores vital product data (VPD) and processor information required for management of CoD features. Because the p5-570 can have processors in up to four physical building blocks, the card can be replaced or updated by an IBM service representative to reflect hardware configuration changes.

> **Note:** In a p5-570 combined system made up of more than one building block, only the two HMC ports and the two service processor communications ports in the building block with the service processor are available to use.

The system supports 32-bit and 64-bit applications, and it requires specific levels of operating system.

2.1.4 Processor features

Each p5-570 building block can contain 2-way processor cards with 64-bit, copper-based, POWER5 microprocessors running at 1.5 GHz, 1.65 GHz, or 1.9 GHz. The processor cards running at 1.5 GHz support up to eight processors per combined system. All card features are available only as Capacity on Demand (CoD). The initial order of the p5-570 system must contain the feature code related to the desired processor card, plus it must contain the processor activation feature code.

2.1.5 Memory features

The processor cards that are used in the p5-570 system offer eight sockets for memory DIMMs. The total memory capacity requires four p5-570 building blocks and eight processor cards. DDR1 and DDR2 DIMMs are different technologies that require different memory sockets, so the processor card with POWER5 microprocessors running at 1.9 GHz is available with two feature codes to allow the two different memory technologies. We recommend that each processor card has an equal amount of memory installed. Balancing memory across the installed processor cards enables memory accesses to be distributed evenly over system components to provide optimal performance.

2.1.6 Disks and media features

Each p5-570 building block features six disk drive bays and two slimline media device bays. In a full configuration with four connected p5-570 building blocks, the combined system supports up to 24 disk bays; therefore, the maximum internal storage capacity is 3.5 TB (using the disk drive features available at the this time of writing). The minimum configuration requires at least one 36.4 GB disk drive. In a full configuration, with four connected p5-570 building blocks, the combined system supports up to eight slimline media device bays. To support two slimline devices in each p5-570 building block, the optional media enclosure and backplane are required.

2.1.7 I/O drawers

The p5-570 has six internal blind swap PCI-X slots: Five are long slots and one is a short slot. The short PCI-X slot can also be used for the Remote I/O expansion card. If more PCI-X slots are needed, such as to extend the number of LPARs and partitions using Micro-Partitioning technology, up to 20 7311 Model D10, 7311 Model D11, or 7311 Model D20 I/O drawers can be attached.

2.1.8 Architecture and technical overview

This section discusses the overall system architecture with its major components described in the following sections. The bandwidths that are provided throughout the section are theoretical maximums used for reference. You should always obtain real-world performance measurements using production workloads.

Memory subsystem

The p5-570 memory controller is internal to the POWER5 chip. It interfaces to either two (DDR1) or four (DDR2) SMI-II buffer chips and eight pluggable DIMMs per processor card. The minimum memory for a p5-570 processor-based system is 2 GB. The maximum installable memory is 512 GB (using DDR1 memory DIMM technology). The p5-570 total memory depends on the number of available processor cards.

PCI-X slots and adapters

PCI-X, where the X stands for extended, is an enhanced PCI bus that delivers a bandwidth of up to 1 GBps, running a 64-bit bus at 133 MHz. PCI-X is backward-compatible, so the p5-570 can support existing 3.3 V PCI adapters.

The system planar provides six PCI-X slots and several integrated PCI devices that interface the three PCI-X to PCI-X bridges to the primary PCI-X buses on the PCI-X host bridge chip.

PCI-X slot 6 can accept a short PCI-X or PCI card, and its space is shared with the Remote I/O expansion card. Therefore, if the Remote I/O expansion card is installed, this slot must remain empty. The remaining PCI-X slots are full-length cards. The dual 1 Gb Ethernet adapter is integrated on the system planar.

The PCI-X slots in the p5-570 system support hot-plug and Extended Error Handling (EEH). In the unlikely event of a problem, EEH-enabled adapters respond to a special data packet that is generated from the affected PCI-X slot hardware by calling system firmware, which examines the affected bus, allows the device driver to reset it, and continues without a system reboot.

64-bit and 32-bit adapters

IBM offers 64-bit adapter options for the p5-570 as well as 32-bit adapters. Higher-speed adapters use 64-bit slots because they can transfer 64 bits of data for each data transfer phase. Generally, 32-bit adapters can function in 64-bit PCI-X slots; however, some 64-bit adapters cannot be used in 32-bit slots. For a full list of the adapters that are supported on the p5-570 systems, and for important information regarding adapter placement, visit the IBM @server® pSeries and AIX Information Center at:

http://publib16.boulder.ibm.com/pseries/en_US/infocenter/base/

Internal storage

Two Ultra320 SCSI controllers under EADS-X (PCI-X host bridge) chips that are integrated into the system planar are used to drive the internal disk drives. The six internal drives plug into the disk drive backplane, which has two separate SCSI buses and controllers with three disk drives per bus. Each of these controllers can be dynamically assigned to partitions if required.

The internal disk drive bays can be used in two different modes, depending on whether the SCSI RAID Enablement Card is installed.

The p5-570 supports a split 6-pack disk drive backplane, which is designed for hot-pluggable disk drives. The disk drive backplane docks directly to the system planar. The virtual SCSI enclosure services (VSES) hot-plug control functions are provided by the Ultra320 SCSI controllers.

> **Note:** Linux does not support hot-swap of any disk drive at the time of writing; therefore, the Linux operating system does not support these hot-swappable procedures. A p5-570 system running Linux must be shut down and powered off before you replace any disk drives.

2.1.9 Dynamic logical partitioning

The logical partition (LPAR) was introduced with the IBM POWER4™ processor product line and the AIX 5L Version 5.1 operating system. The technology offered the capability to divide a system into separate systems, where each LPAR runs an operating environment on dedicated attached devices, such as processors, memory, and I/O components. Clients requested system flexibility to change the system topology on demand, and this was achieved by modifying the system layout on the required HMC. Global or individual changes take part on all involved partitions to redefine the new partition layout. Therefore, a reboot of one or more partitions was required.

Later, dynamic LPAR increased the flexibility, enabling selected system resources such as processors, memory, and I/O components to be added and deleted from dedicated partitions while they were executing. Therefore, AIX 5L V5.2 with all necessary enhancements to enable dynamic LPAR was introduced in 2002. This required an attached HMC with the proper level of software to control the system resources and an updated system firmware level to electronically isolate systems resources. The ability to reconfigure dynamic LPARs encourages system administrators to dynamically redefine all available system resources to reach the optimum capacity for each defined dynamic LPAR.

On the p5-570, the USB devices are considered a group, as are the slimline devices. Devices within a group must be moved from partition to partition as a group. Other devices, such as individual I/O slots, can be relocated individually.

When a system is not featured with the SCSI RAID Enablement Card, each 3-pack of disks will be assigned to an independent integrated SCSI adapter, thus enabling you to configure the system with boot disks for two partitions per building block. When a SCSI RAID Enablement Card is present, all six disks will be logically connected to a single integrated SCSI adapter (even if using JBOD), and therefore, they can only belong to a single active partition at a time.

Virtualization

On the p5-570 server, logical partitions requiring dedicated resources may now be able to take advantage of a new technology that enables resources to be virtualized, allowing for a better overall balance of global system resources and their effective utilization.

Advanced POWER Virtualization feature

The Advanced POWER Virtualization feature is an optional additional cost hardware feature that is available on all IBM eserver POWER5 processor-based systems. Each system has a unique feature code for this feature. The Advanced POWER Virtualization feature includes:

- Firmware enablement for Micro-Partitioning
- Installation image for the IBM Virtual I/O Server software that supports:
  - Ethernet adapter sharing
  - Virtual SCSI server
- Partition Load Manager:
  - Automated CPU and memory reconfiguration
  - Real-time partition configuration and load statistics
- Graphical user interface

For more information about virtualization, go to Chapter 6, “Advanced POWER Virtualization” on page 191.

2.1.10 Boot process

From the earlier RS/6000 systems, through the previous pSeries systems, the boot process passed through several enhancements. With the implementation of the POWER5 technology, the boot process is enhanced to accommodate the flexibility that the POWER5 processor-based hardware features. Depending on the client’s needs, a system might or might not require the use of an HMC to manage the system. The boot process, based on the initial program load (IPL) setup, is determined by the hardware setup and the way you use the features that POWER5 processor-based systems provide.

The IPL process starts when power is connected to the system. Immediately after, the SP starts an internal self test (built-in self-test, or BIST) that is based on integrated diagnostic programs. The system status changes to standby only when all of the test units have passed.

IPL flow without an HMC attached to the system

When system status is standby, the SP presents a System Management Interface (SMI), which can be accessed by pressing any key on an attached serial console keyboard, or the Advanced System Management Interface (ASMI), which uses a Web browser[3] on a client system that is connected to the SP on an

Ethernet network. For more information about ASMI, go to 3.1.1, “Advanced System Management Interface” on page 66.

The SP and the ASMI are standard on all POWER5 processor-based hardware. Both system management interfaces require the general or admin ID password, and they both enable you to set flags that affect the operation of the system according to the provided password, such as auto-power restart, view information about the system (such as the error log and VPD), network environment access setup, and control of system power.

The p5-570 has a permanent firmware boot side, or A side, and a temporary firmware boot side, or B side. New levels of firmware should be installed on the temporary side first in order to test the update’s compatibility with your applications. When the new level of firmware has been approved, it can be copied to the permanent side.

In the SMI and ASMI, you can view and change IPL settings:

- System boot speed

  Fast or Slow: Fast boot results in skipped diagnostic tests and shorter memory tests during the boot.

- Firmware boot side for next boot

  Permanent or Temporary: Firmware updates should be tested by booting from the temporary side before being copied into the permanent side.

- System operating mode

  Manual or Normal: Manual mode overrides various automatic power-on functions, such as auto-power restart, and enables the power switch button.

- AIX 5L/Linux partition-mode boot (available only if the system is not managed by the HMC)
  - Service mode boot from saved list: This is the preferred way to run concurrent AIX diagnostics.
  - Service mode boot from default list: This is the preferred way to run stand-alone AIX 5L diagnostics.
  - Boot to open firmware prompt.
  - Boot to System Management Service (SMS): To further select the boot devices or network boot options.
- Boot to server firmware
  - Select the state for the server firmware: Standby or Running.

[3] Supported browsers are Netscape (Version 7.1), Microsoft® Internet Explorer (Version 6.0), and Opera (Version 7.23). At the time of writing, older or previous versions of these browsers are not supported. JavaScript™ and cookies must be enabled.

  - When the server is in the server firmware standby state, partitions can be set up and activated.

IPL flow with an HMC attached to the system

When system status is standby, you can either use the HMC to open a virtual terminal and access the SMI, or launch a Web browser to access the ASMI.

Using the SMI or the ASMI, you can view or modify the proper IPL settings in order to set the boot mode to partition standby and then turn the system on, but the HMC can be also used to power on the managed system (and is highly recommended). Using the HMC to turn the system on requires selecting one of the following options:

- Partition Standby

  The Partition Standby power-on mode enables you to create and activate logical partitions:

  - When the Partition Standby power-on is completed, the operator panel on the managed system displays `LPAR...`, indicating that the managed system is ready for you to use the HMC to partition its resources and, possibly, activate them.
  - When a partition is activated, the HMC requires you to select the boot mode of the single partition.

- System Profile

  The System Profile option powers on the system according to a predefined set of profiles. Profiles are activated in the order in which they are shown in the system profile.

- Partition autostart

  This option powers on the managed system to partition standby mode and then activates all partitions that have been designated autostart.

After the system succeeds to boot with any of these choices, the HMC can be used to manage the system, such as continuing to boot from operating system or managing the logical partitions.

> **Note:** The same HMC cannot be attached to POWER4 and POWER5 processor-based systems simultaneously, but for redundancy purposes, one POWER5 server can be attached to two HMCs.

2.1.11 Operating system requirements

All new POWER5 servers are capable of running IBM AIX 5L for POWER and support appropriate versions of Linux. AIX 5L has been specifically developed and enhanced to exploit and support the extensive RAS features on IBM @server pSeries systems.

For more information about the supported operating systems, go to 3.5.1, “Operating system capability and configuration” on page 92.

2.1.12 Capacity on Demand

p5-570 systems can be shipped with non-activated resources (processors and memory) that can be added as they are needed. Processors and memory can be brought online to meet increasing workload demands without affecting system operations.

IBM @server Capacity on Demand (CoD) is supported on p5 Model 570 when the 1.65 GHz and 1.9 GHz POWER5 processors are used.

The following sections outline the different methods that are available, namely:

- Processor
  - Capacity Upgrade on Demand
  - Reserve Capacity on Demand
  - Dynamic processor sparing
- Memory
  - Permanent Capacity Upgrade on Demand for memory
  - On/Off Capacity on Demand
  - Reserve Capacity on Demand
- Processor and memory
  - Trial Capacity on Demand for processors and memory

2.2 p5-575

The 9118 IBM @server p5 Model 575 delivers an 8-way, 1.9 GHz POWER5 high-bandwidth cluster node, ideal for many high-performance computing applications.

The p5-575 offers a highly competitive platform for addressing the needs of users who can take advantage of the distinctive combination of high-performance computation and extraordinary per-processor memory bandwidth offered by the p5-575. The p5-575 is designed with a strong affinity to applications such as oceanographic and atmospheric studies, quantum physics, and computer-aided engineering (CAE), as well as meeting the requirements of supporting large-scale, data-intensive applications, such as business intelligence and data warehousing found in the retail, banking, finance, and insurance sectors, among others.

The p5-575 is packaged in a super-dense 2U form factor, with up to 12 nodes installed in a 42U-tall, 24-inch rack. Multiple racks of p5-575 nodes can be combined to provide a broad range of powerful cluster solutions. Up to 16 p5-575 nodes can be clustered together for a total of 128 processors. Clusters larger than 16 nodes are available through special bid at this time; clusters of 128 nodes (1024 processors) are available.

The symmetric multiprocessor (SMP) node uses 64-bit, copper-based, single-core POWER5 microprocessors in an 8-way configuration. Each microprocessor is supported by 36 MB of Level 3 cache and up to eight memory DIMMs with point-to-point memory to processor connections. Each 8-way includes 64 slots for memory DIMMs. Memory sizes are offered from 1 GB up to 256 GB. With the optional I/O Assembly with PCI-X and RIO-2, along with a system I/O drawer, up to twenty-four PCI-X cards and up to eighteen 15,000 rpm disk drives are available.

The p5-575 offers:

- 2U rack-mount design
- 1.9 GHz single-core IBM POWER5 processors in 8-way configuration
- 12.4 GBps memory bandwidth per CPU
- Up to 256 GB of memory
- Up to 1,468 GB of internal disk storage
- Up to 18 hot-swap disk bays and 24 hot-swap PCI-X slots
- Up to six independent I/O buses
- 36 MB of low-latency cache per CPU
- Redundant rack power subsystem
- Dynamic logical partitioning (DLPAR)
- Optional Advanced POWER Virtualization
- Cluster 1600 enhancements include Cluster Systems Management (CSM) support for the p5-575

2.2.1 System specifications

Table 2-3 lists the general system specifications of a p5-575 system.

Table 2-3 p5-575 specifications

Description	Range
Operating temperature	5 to 32 degrees C (41 to 95° F)
Relative humidity	8% to 80%
Maximum wet bulb	23 degrees C (73° F) (operating)
Operating voltage	200 to 240 V AC 50/60 Hz
Maximum power consumption	41.6 kW (maximum)
Maximum power source loading	22.3 (maximum, assuming 0.93 pf) kVA (x2)
Maximum thermal output	41,600 joules/sec (142,000 Btu/hr) maximum

2.2.2 Physical package

The p5-575 cluster node uses the 24-inch System Rack (FC 5793) with 42U of rack space for:

- p5-575 nodes
- Two identical, fully redundant bulk power assemblies (BPA)
- Two identical, fully redundant, optional internal battery features (IBF)
- I/O drawers
- Covers, front and rear (option for either slimline or acoustic)

The p5-575 rack incorporates two identical, redundant BPAs mounted in the front and rear (redundant) sections of the top of the rack. The rack has an option for fully redundant internal battery features (IBF) mounted in the front and rear (redundant) sections of the rack, as shown in Table 2-4.

Table 2-4 Physical packaging of the p5-575 rack

Dimension	p5-575
Height	2025 mm (79.7 in.)
Width	785 mm (30.9 in.)
Depth	1530 mm (60.2 in.)
Weight	1573 kg (3460 lb) with acoustic door and IBF

2.2.3 Minimum system configuration

The minimum configuration for an IBM eserver 9118 575 includes the items (for the first node in a rack) listed in Table 2-5.

Table 2-5 Minimum system configuration

Quantity	Component description
One	p5-575 (9118-575)
One	8-way 1.9 GHz POWER5 processor, 288 MB L3 cache and system planar
One	I/O assembly without PCI-X or RIO-2 capability
One	Power cable group, 09U (or 09U Left)
One	Ethernet cable, node to hubs, EIA01 through EIA17
One	1024 MB (4 x 256 MB) DIMMs, 208-pin, 266 MHz DDR SDRAM
Two	73.4 GB 10,000 rpm Ultra320 SCSI disk drive assembly
One	System rack
One	Rack content specify
Two	Slimline or acoustic door kit
Two	Bulk power regulators
Two	Power controller assemblies
Two	Line cords
One	Ethernet cable, Hardware Management Console to system unit
One	Language specify

2.2.4 Memory

The p5-575 uses Double Data Rate (DDR) DRAM memory DIMMs.

2.2.5 Logical partitioning

Logical partitioning (LPAR) allows the p5-575 node resources to be allocated and for multiple instances of the supported operating systems to be run simultaneously on a single node:

- LPAR allocation, monitoring, and control is provided by the Hardware Management Console.

- Each LPAR functions under its own instance of the operating system.
- A minimum of 128 MB of memory is required per LPAR.

2.2.6 Advanced POWER Virtualization and Partition Load Manager

The optional Advanced POWER Virtualization feature allows partitions to be created that are in units of less than one CPU (sub-CPU LPARs) and allows the same system I/O to be virtually allocated to these partitions. For more details, go to Chapter 6, “Advanced POWER Virtualization” on page 191.

- With Advanced POWER Virtualization, the processors on the system can be partitioned into as many as 10 LPARs per processor.
- Advanced POWER Virtualization includes Partition Load Manager, which provides cross-partition workload management across the system LPARs.
- An encryption key is supplied to the client and installed on the system, authorizing the partitioning at the subprocessor level.
- Using Advanced POWER Visualization, the p5-575 can be divided into as many as 80 logical partitions per node. System resources can be dedicated to each LPAR.
- Advanced POWER Virtualization requires AIX 5L V5.3 or SUSE LINUX Enterprise Server 9 for POWER.

2.2.7 System control

The following key points pertain to system control:

- Each p5-575 node must be connected to a Hardware Management Console (HMC) for system control, LPAR, and service functions. The HMC is capable of supporting multiple POWER5 nodes.
- Each p5-575 node can be connected to two HMCs for redundancy if desired.
- The p5-575 is connected to the HMC using Ethernet connections. Each P5-575 connects to the integrated Ethernet hubs within the system rack. The hubs are connected to the HMC through two Ethernet cables.

2.2.8 I/O drawers

The following key points pertain to available I/O drawers:

- The p5-575 uses optional 4U-tall remote I/O drawers for additional directly attached PCI or PCI-X adapters and SCSI disk capabilities.

- Each I/O drawer is divided into halves. Each half contains 10 blind-swap PCI-X slots and one or two Ultra3 SCSI 4-pack backplanes for a total of 20 PCI slots and up to 16 hot-swap disk bays per drawer.
- Existing 7040-61D I/O drawers can be attached to a p5-575 node if available.
- Only 7040-61D I/O drawers containing PCI-X planars are supported. Any PCI planars must be replaced with PCI-X planars before the drawer can be attached.
- SP Switch2 PCI-X Attachment Adapter must be removed from the 7040-61D if present.
- A maximum of one I/O drawer can be connected to a p5-575 node.
- One single-wide, blind-swap cassette is provided in each PCI-X slot of the I/O drawer. To ensure the proper environmental characteristics for the drawer, FC 4599 should be ordered for adapters requiring a blind swap cassette.
- All 10 PCI-X slots on each I/O drawer planar are capable of supporting either 64-bit or 32-bit PCI or PCI-X adapters. Each I/O drawer planar provides 10 PCI-X slots capable of supporting 3.3 V signaling PCI or PCI-X adapters operating at speeds up to 133 MHz.
- Each I/O drawer planar incorporates two integrated Ultra3 SCSI adapters for direct attachment of the two 4-pack hot swap backplanes in that half of the drawer. These adapters do not support external SCSI device attachments.
- Each half of the I/O drawer is powered separately.

2.2.9 Disks, boot devices, and media devices

A minimum of two identical internal SCSI hard disk drives are required per p5-575 node. We highly recommend that these disks be used as mirrored boot devices. This configuration provides service personnel the maximum amount of diagnostic information if the system encounters errors in the boot sequence.

The p5-575 incorporates an early power-off warning (EPOW) capability that assists in performing an orderly system shutdown in the event of a sudden power loss. IBM recommends use of the Integrated Battery Backup features or an uninterruptible power system (UPS) to help ensure against loss of data due to power failures.

2.2.10 PCI-X slots and adapters

System maximum limits for adapters and devices might not provide optimal system performance. These limits are given for connectivity and functionality assurance.

Configuration limitations have been established to help ensure appropriate PCI or PCI-X bus loading, adapter addressing, and system and adapter functional characteristics when ordering I/O drawers.

Adapters

Most PCI and PCI-X adapters for the p5-575 system are capable of being hot-plugged. Any PCI adapter supporting a boot device or system console should not be hot-plugged.

The following adapters are not hot-plug-capable:

- POWER GXT135P Graphics Accelerator with Digital Support
- 2-Port Multiprotocol PCI Adapter

2.2.11 Software requirements

The minimum software requirements are:

- AIX 5L for POWER V5.2 with the 5200-04 Recommended Maintenance Package (APAR IY56722), or later, plus APAR IY60347
- AIX 5L for POWER V5.3 with APAR IY60349, or later
- SUSE LINUX Enterprise Server 9 for POWER systems, or later
- Red Hat Enterprise Linux AS for POWER Version 3, or later

For more information about software requirements, see 3.5.1, “Operating system capability and configuration” on page 92.

2.2.12 For the required Hardware Management Console

The HMC must have Hardware Management Console for POWER5 Licensed Machine Code Version 4.4 provided in APAR MB00691.

2.3 p5-590 and p5-595

The IBM eserver p5 590 and IBM eserver p5 595 servers are redefining the IT economics of enterprise UNIX and Linux computing. The up to 64-way p5-595 server is the new flagship of the product line with nearly three times the commercial performance (based on rPerf estimates) and twice the capacity of its predecessor, the IBM eserver pSeries 690. Accompanying the p5-595 is the up to 32-way p5-590 that offers enterprise-class function and more performance than the pSeries 690 at a significantly lower price for comparable configurations.

These servers come standard with mainframe-inspired reliability, availability, and serviceability (RAS) capabilities and IBM Virtualization Engine™ systems technology with breakthrough innovations such as Micro-Partitioning. Micro-Partitioning allows as many as 10 logical partitions (LPARs) per processor to be defined. Both systems can be configured with up to 254 virtual servers with a choice of the AIX 5L, Linux, and i5/OS operating systems in a single server, designed to enable cost-saving consolidation opportunities.

> **Note:** Not all system features available under the AIX 5L operating system are available under the Linux operating system.

2.3.1 Model abstract for 9119-590 and 9119-595

The 9119 IBM @server p5 Model 590 and 595 systems provide an expandable high-end enterprise solution for managing e-business computing requirements.

Table 2-6 represents the major product attributes of these models with the major differences highlighted by shading.

Table 2-6 9119-590 and 9119-595 attributes

Attribute	9119-590	9119-595
SMP processor configurations	8- to 32-way	16-, 32-, 48- and 64-way
Maximum 16-way CPU books	2	4
Processor clock rate	1.65 GHz	1.65 GHz Standard or 1.9 GHz Turbo
Processor cache per processor pair	1.9 MB Level 2 36 MB Level 3	1.9 MB Level 2 36 MB Level 3
Processor packaging	Multichip Module (MCM)	MCM
64-bit copper technology POWER5 processor	Y	Y
Maximum memory configuration	1 TB	2 TB
Rack space	42U 24-inch custom rack	42U 24-inch custom rack
Maximum number of I/O drawers	8	12
Maximum number of PCI-slots	160	240

Attribute	9119-590	9119-595
Maximum number of 15 K rpm disks	128	192
Dual service processors	Y	Y
Integrated redundant power	Y	Y
Battery backup option	Y	Y
Powered expansion rack available	N	Y
Dynamic LPAR	Y	Y
Micro-Partitioning with up to 254 partitions	Y	Y
Acoustic rack doors available	Y	Y
Support for AIX 5L, Linux, and i5/OS	Y	Y

Each 16-way processor book also includes 16 slots for memory cards and six Remote I/O-2 attachment cards for connection of the system I/O drawers.

Each I/O drawer contains twenty 3.3-volt PCI-X adapter slots and up to sixteen disk bays.

The AIX 5L V5.2 and V5.3, Linux, and i5/OS V5R3 operating systems can run simultaneously in different partitions within the a server.

2.3.2 System frames

Both the p5-590 and p5-595 systems are based on the same 24-inch wide, 42 EIA height frame. Inside this frame, all the server components are placed in predetermined positions. This design and mechanical organization offers advantages in optimization of floor space usage.

The p5-590 and p5-595 servers are designed with a basic server configuration that starts with a single *frame* (Figure 2-2 on page 28) and is featured with optional and required components.

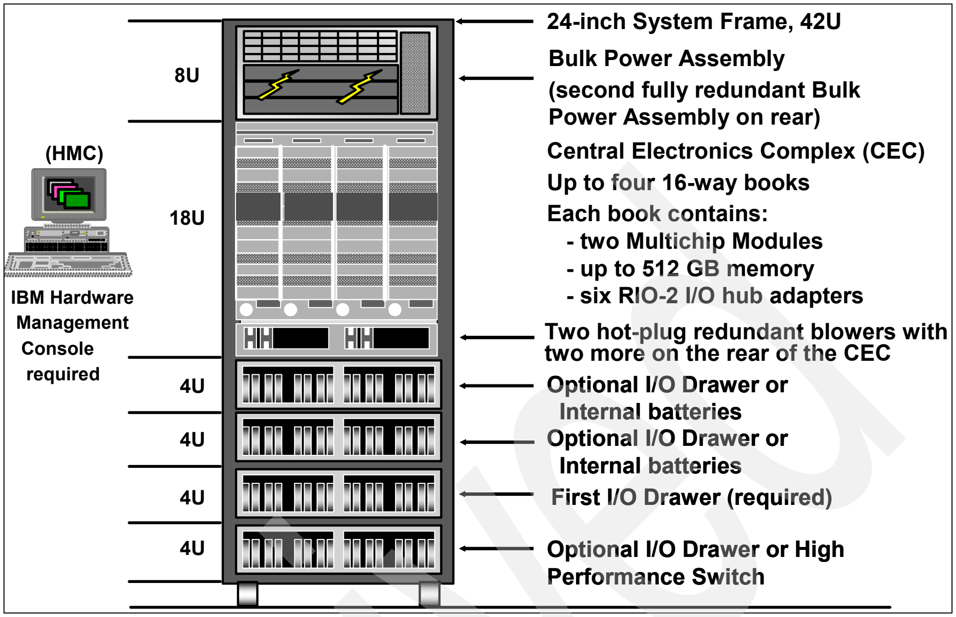
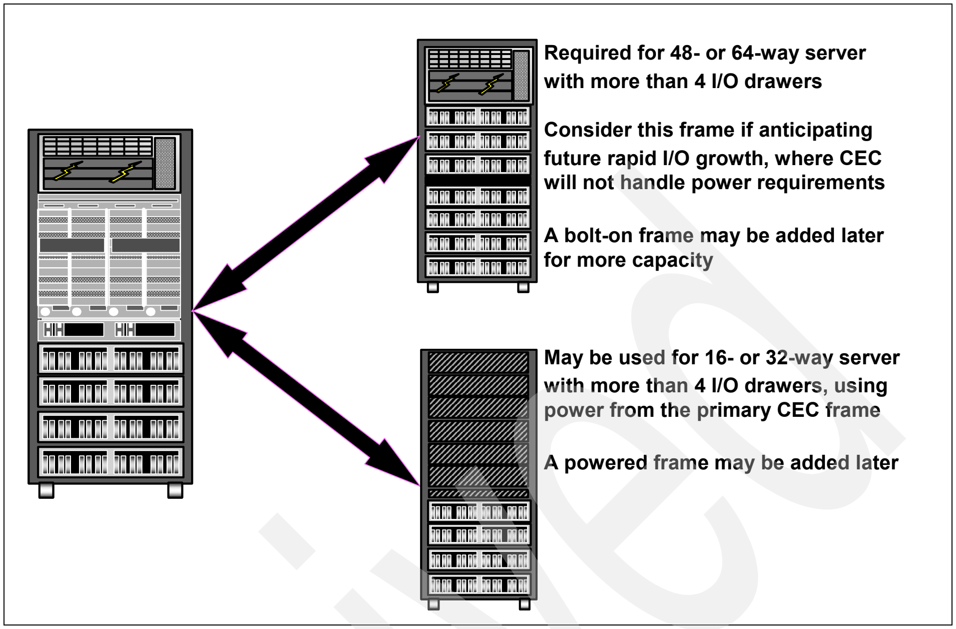

8U
18U
4U
4U
4U
4U
(HMC)
IBM Hardware Management Console required
24-inch System Frame, 42U
Bulk Power Assembly
(second fully redundant Bulk Power Assembly on rear)
Central Electronics Complex (CEC)
Up to four 16-way books
Each book contains:
- two Multichip Modules
- up to 512 GB memory
- six RIO-2 I/O hub adapters
Two hot-plug redundant blowers with two more on the rear of the CEC
Optional I/O Drawer or Internal batteries
Optional I/O Drawer or Internal batteries
First I/O Drawer (required)
Optional I/O Drawer or High Performance Switch

Figure 2-2 Primary system frame organization

For additional capacity, either a powered or non-powered frame can be configured for a p5-595, or a non-powered frame for the p5-590, as shown in Figure 2-3 on page 29.

A powered Expansion Rack is available for large system configurations that require more power and space than is available from the primary system rack. It provides the same redundant power subsystem available in the primary rack. For example, when p5-595 servers contain three or four processor books, the power subsystem in the primary system rack can support only the central electronics complex (CEC) and any I/O drawers that can be mounted in the system rack itself. In such configurations, additional I/O drawers must be mounted in the powered Expansion Rack.



Required for 48- or 64-way server with more than 4 I/O drawers
Consider this frame if anticipating future rapid I/O growth, where CEC will not handle power requirements
A bolt-on frame may be added later for more capacity
May be used for 16- or 32-way server with more than 4 I/O drawers, using power from the primary CEC frame
A powered frame may be added later

Figure 2-3 Powered and bolt-on frames

2.3.3 Installation planning

Product installation and in-depth system cabling are beyond the scope of this publication. Complete installation instructions are shipped with each order. Comprehensive planning advice is available at the following Web page:

```
http://publib.boulder.ibm.com/infocenter/pseries/index.jsp
```

We describe key specifications in the following sections.

System specifications

Table 2-7 lists the general system specifications of the p5-590 and p5-595 servers.

Table 2-7 p5-590 and p5-595 server specifications

Description	Range
Recommended operating temperature (8-way, 16-way, 32-way)	10 degrees to 32 degrees C (50 degrees to 89.6 degrees F)
Recommended operating temperature (48-way and 64-way)	10 degrees to 28 degrees C (50 degrees to 82.4 degrees F)

Description	Range
Operating voltage	200 to 240, 380 to 415, or 480 volts AC
Operating frequency	50/60 plus or minus 0.5 Hz
Maximum power consumption (1.9 GHz processor)	22.7 kW
Maximum power consumption (1.65 GHz processor)	20.3 kW
Maximum thermal output (1.9 GHz processor)	77.5 kBtu/hr (British thermal unit)
Maximum thermal output (1.65 GHz processor)	69.3 kBtu/hr (British thermal unit)

Physical package

Table 2-8 lists the major physical attributes found on the p5-590 and p5-595 servers.

Table 2-8 p5-590 and p5-595 server physical packaging

Dimension	
Height	2025 mm (79.7 in.)
Width	785 mm (30.9 in.)
Depth	1326 mm (52.2 in.)[a] or 1681 mm (66.2 in.)[b]
Weight	
Minimum configuration	1241 kg (2735 lb)
Maximum configuration	2458 kg (5420 lb)

a. With slimline doors installed
b. Without slimline doors installed

There are several possible frame configurations of the p5-590 and p5-595 servers. FC 7960 (Compact Rack Option) is designed to provide improved access through doorways during shipment.

Figure 2-4 on page 31 shows service clearances for double-frame systems with acoustical doors.

> **Note:** The p5-595 and p5-590 servers must be installed in a raised floor environment.

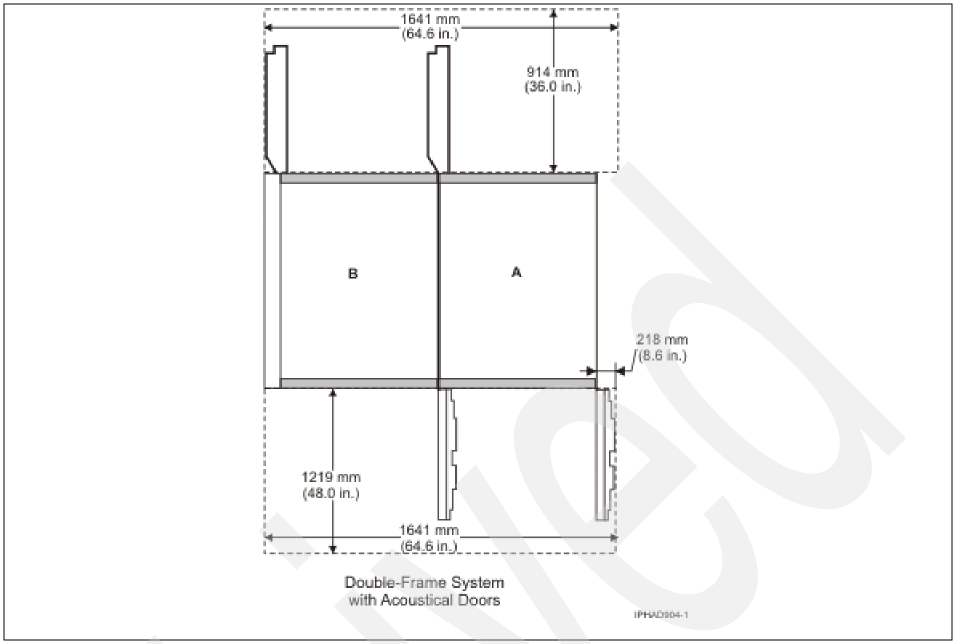


Figure 2-4 Service clearances

2.3.4 Minimum and optional features

The purpose of this section is to establish the minimum configuration for a p5-590 and p5-595.

Table 2-9 identifies the components required to construct a minimum configuration for a p5-590.

Table 2-9 p5-590 minimum system configuration

Quantity	Component description
One	IBM @server p5-590.
One	Media drawer for installation and service actions (additional media features might be required) without Network Installation Management (NIM).
One	16-way, POWER5 processor book, 0-way active.
Eight	1-way, processor activations.
Two	Memory cards with a minimum of 8 GB of activated memory.
Two	Processor clock cards, programmable.

Quantity	Component description
One	Power cable group, bulk power to CEC and fans.
Three	Power converter assemblies, central electronics complex.
One	Power cable group, first processor book.
Two	System service processors.
One	Multiplexer card.
Two	RIO-2 loop adapters, single loop.
One	I/O drawer. Note: Requires 4U frame space.
One	Remote I/O (RIO) cable, 0.6 m. Note: Used to connect drawer halves.
Two	Remote I/O (RIO) cables, 2.5 m.
Two	15,000 rpm Ultra3 SCSI disk drive assemblies.
One	I/O drawer attachment cable group.
One	Slimline or acoustic door kit.
Two	Bulk power regulators.
Two	Bulk Power Controller assemblies.
Two	Bulk power distribution assemblies.
Two	Line cords.
One	Language specify.
One	Hardware Management Console.

Table 2-10 identifies the components required to construct a minimum configuration for a p5-595.

Table 2-10 p5-595 minimum system configuration

Quantity	Component description
One	IBM @server p5-595.
One	16-way, POWER5 processor book, 0-way active.

Quantity	Component description
Note: The following two components must be added to p5-595 servers with one processor book: ▶ One - Cooling group ▶ One - Power cable group	
Sixteen	1-way, processor activations.
Two	Memory cards with a minimum of 8 GB of activated memory.
Two	Processor clock cards, programmable.
One	Power cable group, bulk power to CEC and fans.
Three	Power converter assemblies, central electronics complex.
One	Power cable group, first processor book.
One	Multiplexer card.
Two	Service processors.
Two	RIO-2 loop adapter, single loop.
One	I/O drawer. Note: 4U frame space required.
One	Remote I/O (RIO) cable, 0.6 m. Note: Used to connect drawer halves.
Two	Remote I/O (RIO) cables, 3.5 m.
Two	15,000 rpm Ultra3 SCSI disk drive assembly.
One	PCI SCSI Adapter or PCI LAN Adapter for attachment of a device to read CD media or attachment to a NIM server.
One	I/O drawer attachment cable group.
One	Slimline or acoustic door kit.
Two	Bulk power regulators.
Two	Power controller assemblies.
Two	Power distribution assemblies.
Two	Line cords.
One	Language specify.
One	Hardware Management Console.

Note: An HMC is required, and two HMCs are recommended. A private network with the HMC providing DHCP services is mandatory on these systems.

2.3.5 Processor features

The p5-590 system features base 8-way Capacity on Demand (CoD), 16-way, and 32-way configurations with the POWER5 processor running at 1.65 GHz. The p5-595 system features base 16-way, 32-way, 48-way, and 64-way configurations with the POWER5 processor running at 1.65 GHz or 1.9 GHz.

The p5-590 and p5-595 system configuration is based on the processor book. To configure it, it is necessary to order one or more of the following components:

- One or more 16-way processor book, 0-way active
- Activation codes to reach the expected configuration

Note: Any p5-595 or p5-590 system made up of more than one processor book must have all processor cards running at the same speed.

For a list of available processor features, refer to Table 2-11.

Table 2-11 Available processor options

Description
16-way POWER5 Turbo Capacity Upgrade on Demand (CUoD) processor book, 0-way active
16-way POWER5 Standard Capacity Upgrade on Demand (CUoD) processor book, 0-way active

Note: POWER5 Turbo refers to 1.9 GHz clocking; POWER5 Standard refers to 1.65 GHz clocking.

2.3.6 Operating system support

The p5-590 and p5-595 server are capable of running IBM AIX 5L for POWER and i5/OS and support appropriate versions of Linux. AIX 5L has been specifically developed and enhanced to exploit and support the extensive RAS features on IBM @server systems.

For more details, see 3.5.1, “Operating system capability and configuration” on page 92.

2.3.7 Memory subsystem

The p5-590 and p5-595 memory controller is internal to the POWER5 chip. It interfaces to four Synchronous Memory Interface II (SMI-II) buffer chips and eight DIMM cards per processor chip. There are 16 memory card slots per processor book, and each processor chip on an Multichip Module (MCM) owns a pair of memory cards.

The p5-590 and p5-595 use Double Data Rate (DDR) DRAM memory cards. The two types of DDR memory used are DDR1 and the higher-speed DDR2. Memory migration from previous systems is not supported.

> **Note:** Because the DDR1 and DDR2 modules use different voltages, mixing the memory technologies is not allowed within a server.

Table 2-12 lists the memory features available for the p5-590 and the p5-595 at the time of writing.

Table 2-12 Types of available memory cards for p5-590 and p5-595

Memory type	Size	Speed	Number of memory cards
DDR1 COD	4 GB (2 GB active)	266 MHz	1
	8 GB (4 GB active)	266 MHz	1
DDR1	16 GB	266 MHz	1
	32 GB	200 MHz	1
	256 GB package	266 MHz	32 * 8 GB
	512 GB package	266 MHz	32 * 16 GB
	512 GB package	200 MHz	16 * 32 GB
DDR2	4 GB	533 MHz	1

Memory configuration and placement

The minimum memory for a p5-590 processor-based system is 2 GB, and the maximum installable memory is 1024 GB using DDR1 memory DIMM technology (128 GB using DDR2 memory DIMM). The total memory depends on the number of available processors (16 per processor book).

Table 2-13 lists the possible memory configurations.

Table 2-13 Memory configuration table

System	p5-590	p5-595
Min. configurable memory	8 GB	8 GB
Max. configurable memory using DDR1 memory	1,024 GB	2,048 GB
Max. configurable memory using DDR2 memory	128 GB	256 GB
Max. number of memory cards	32	64

The following rules *must* be observed:

- Memory must be installed in identical pairs.
- Servers with one processor book must have a minimum of two memory cards installed.
- Servers with two processor books must have a minimum of four memory cards installed per processor book (two per MCM).

The following memory configuration guidelines are *recommended*:

- The same amount of memory should be used for each MCM (two per processor book) in the system.
- Each 8-way MCM (two per processor book) should have some memory.
- No more than two different sizes of memory cards should be used in each processor book.
- All MCMs (two per processor book) in the system should have the same aggregate memory size.
- A minimum of half of the available memory slots in the system should contain memory.
- It is better to install more cards of smaller capacity than fewer cards of larger capacity.

For p5-590 and p5-595 servers being used for high-performance computing, the following guidelines are *strongly recommended*:

- Use DDR2 memory.
- Install some memory in support of each 8-way MCM (two MCMs per processor book).
- Use the same sized memory cards across all MCMs and processor books in the system.

2.3.8 Internal I/O subsystem

The p5-590 and p5-595 use remote I/O drawers (that are 4U) for directly attached PCI or PCI-X adapters and SCSI disk capabilities. A minimum of one I/O drawer is required per system.

> **Note:** The p5-590 supports up to eight I/O drawers, while the p5-595 supports up to 12 I/O drawers.

2.3.9 Disks and boot devices

A minimum of two internal SCSI hard disks are required per server. We recommend that these disks be used as mirrored boot devices. These disks should be mounted in the first I/O drawer whenever possible. This configuration provides service personnel the maximum amount of diagnostic information if the system encounters errors in the boot sequence.

Boot support is also available from local SCSI, SSA, and Fibre Channel adapters, or from networks using Ethernet or token-ring adapters.

If the boot source other than internal disk is configured, the supporting adapter should also be in the first I/O drawer.

2.3.10 Media options

The p5-590 and p5-595 servers must have access either to a device capable of reading CD media or to a Network Installation Management (NIM) server:

- The recommended devices for reading CD media are the rack-mounted media drawer or an IBM Storage Device Enclosure. The media drawer is mounted in the 13U location of the CEC Rack; the 7212-102, 7210-025, and 7210-030 enclosures attach using a PCI SCSI adapter in one of the system I/O drawers.
- If a NIM server is used, it must attach through a PCI LAN adapter in one of the system I/O drawers. We recommend an Ethernet connection.

The rack-mounted media drawer provides a 1U high internal media drawer for use with the p5-590 and p5-595 servers. The media drawer displaces any I/O drawer or battery backup feature components that would be located in the same location of the primary system rack. The media drawer provides a fixed configuration that must be ordered with three media devices and all required SCSI and power attachment cabling.

This media drawer can be mounted in the CEC rack with three available media bays, two in the front and one in the rear. The device in the rear is only accessible from the rear of the system. New storage devices for the media bays include:

- 16X/48X IDE DVD-ROM drive
- 4.7 GB, SCSI DVD-RAM drive
- 36/72 GB, 4 mm internal tape drive

2.3.11 PCI-X slots and adapters

PCI-X, where the X stands for extended, is an enhanced PCI bus, delivering a theoretical peak bandwidth of up to 1 GBps, running a 64-bit bus at 133 MHz. PCI-X is backward compatible, so the p5-590 and p5-595 I/O drawers can support existing 3.3 volt PCI adapters.

Most PCI and PCI-X adapters for the p5-590 and p5-595 servers are capable of being hot-plugged. Any PCI adapter supporting a boot device or system console should not be hot-plugged. The following adapters are not hot-plug-capable:

- POWER GXT135P Graphics Accelerator with Digital Support
- 2-Port Multiprotocol PCI Adapter

System maximum limits for adapters and devices might not provide the optimal system performance. These limits are given to help assure connectivity and function.

Configuration limitations have been established to help ensure appropriate PCI or PCI-X bus loading, adapter addressing, and system and adapter functional characteristics when ordering I/O drawers. These I/O drawer limitations are in addition to individual adapter limitations shown in the feature descriptions section of the *IBM Universal Sales Manual* CD-ROM (SK3T-8287-59).

The maximum number of a specific PCI or PCI-X adapter allowed per p5-590 and p5-595 servers might be less than the number allowed per I/O drawer multiplied by the maximum number of I/O drawers.

The PCI-X slots in the I/O drawers of p5-590 and p5-595 servers support Extended Error Handling (EEH). In the unlikely event of a problem, EEH-enabled adapters respond to a special data packet generated from the affected PCI-X slot hardware by calling system firmware, which is designed to examine the affected bus, allow the device driver to reset it, and continue without a system reboot.

Note: As soon as a p5-590 or p5-595 server is connected to a Hardware Management Console, the POWER Hypervisor will prevent the system from using non-EEH OEM adapters.

2.3.12 Internal storage

Each I/O drawer contains four integrated Ultra3 SCSI adapters and SCSI Enclosure Services (SES hot-swappable control functions).

Each of the 4-packs supports up to four hot-swappable Ultra3 SCSI disk drives, which can be used for the installation of the operating system or storing data.

Table 2-14 lists hot-swappable disk drives.

Table 2-14 Hot-swappable disk drive options

Description
36.4 GB 15,000 rpm Ultra320 SCSI hot-swappable disk drive
73.4 GB 15,000 rpm Ultra320 SCSI hot-swappable disk drive

Note: Disks with 10,000 rpm rotational speeds from earlier systems are not supported.

Prior to the hot-swap of a disk in the hot-swap-capable bay, all necessary operating system actions must be undertaken to ensure that the disk is capable of being deconfigured. After the disk drive has been deconfigured, the SCSI enclosure device will power off the bay, enabling the safe removal of the disk. You should ensure that the appropriate planning has been given to any operating-system-related disk layout, such as the AIX 5L Logical Volume Manager, when using disk hot-swap capabilities. For more information, see *Problem Solving and Troubleshooting in AIX 5L*, SG24-5496.

2.4 Simultaneous multithreading

Simultaneous multithreading is the ability of a single physical processor to simultaneously dispatch instructions from more than one hardware thread context. Because there are two hardware threads per physical processor, additional instructions can run at the same time. The POWER5 processor is a superscalar processor that is optimized to read and run instructions in parallel. Simultaneous multithreading enables you to take advantage of the superscalar nature of the POWER5 processor by scheduling two applications at the same time on the same processor. No single application can fully saturate the processor. Simultaneous multithreading is a feature of the POWER5 processor and is available with shared processors.

The simultaneous multithreading mode maximizes the usage of the execution units. In the POWER5 chip, more rename registers have been introduced (for floating-point operation, rename registers increased to 120), which are essential for out of order execution, and then vital for simultaneous multithreading.

Figure 2-5 shows two threads executed on the same processor.

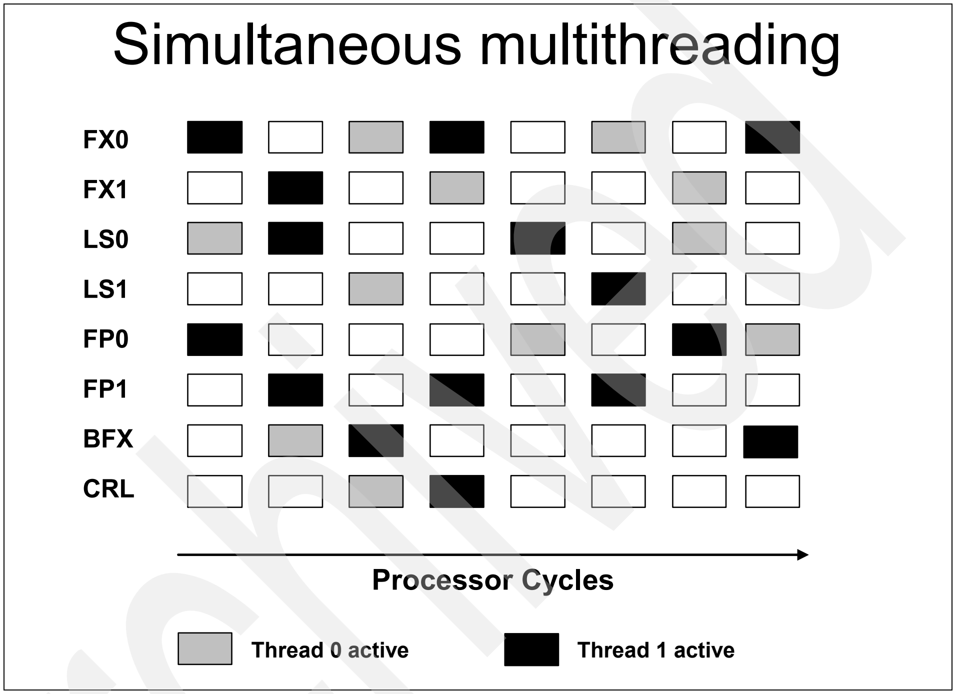


Figure 2-5 Simultameous multithreading

If simultaneous multithreading is activated:

- More instructions can be executed at the same time.
- A single physical POWER5 processor appears to the operating system to be two logical processors.
- Support is provided in mixed environments:
  - Capped and uncapped partitions
  - Virtual partitions
  - Dedicated partitions
  - Single partition systems

Note: Simultaneous multithreading is supported on POWER5 processor-based systems running AIX 5L Version 5.3 or Linux operating system-based systems at an appropriate level. AIX 5L Version 5.2 does not support this function.

The simultaneous multithreading policy is controlled by the operating system and is thus partition specific. AIX 5L provides the **`smtctl`** command that turns simultaneous multithreading on and off either immediately, or on next reboot. For a complete listing of flags, see:

http://publib.boulder.ibm.com/infocenter/pseries/index.jsp

For Linux, an additional boot option must be set to activate simultaneous multithreading after a reboot.

2.4.1 Enhanced simultaneous multithreading features

To improve simultaneous multithreading performance for various workloads and provide robust quality of service, the POWER5 processor provides two features:

- Dynamic resource balancing

  Dynamic resource balancing is designed to ensure that the two threads executing on the same processor flow smoothly through the system. Depending on the situation, the POWER5 processor resource balancing logic has different thread throttling mechanisms (a thread that reaches a threshold of L2 cache misses will be throttled to allow other threads to pass the stalled thread).

- Adjustable thread priority

  Adjustable thread priority that allows software to determine when one thread should have a greater (or lesser) share of execution resources. The POWER5 processor supports eight software-controlled priority levels for each thread.

2.4.2 Single threading operation

Having threads executing on the same processor will not increase the performance of applications with execution unit limited performance, or applications that consume all the chip's memory bandwidth. For this reason, the POWER5 processor supports the single threading execution mode. In this mode, the POWER5 processor gives all the physical resources to the active thread, allowing it to achieve higher performance than a POWER4 processor based-system at equivalent frequencies. Highly optimized scientific codes are one example where single threading operation may provide more throughput.

2.4.3 Benefits of simultaneous multithreading

To provide improved performance at the application level, simultaneous multithreading functionality is embedded in the POWER5 chip technology. Applications developed to use process-level parallelism (multitasking) and thread-level parallelism (multithreads) can shorten their overall execution time. Simultaneous multithreading is the next stage of processor saturation for throughput-oriented applications to introduce the method of instruction-level parallelism to support multiple pipelines to the processor.

Simultaneous multithreading is primarily beneficial in commercial environments where the speed of an individual transaction is not as important as the total number of transactions that are performed. Simultaneous multithreading is expected to increase the throughput of workloads with large or frequently changing working sets, such as database servers and Web servers.

Workloads that see the greatest simultaneous multithreading benefit are those that have a high cycles per instruction (CPI) count. These workloads tend to use processor and memory resources poorly. Large CPIs are usually caused by high cache-miss rates from a large working set. Large commercial workloads are somewhat dependent on whether the two hardware threads share instructions or data, or the hardware threads are completely distinct.

Workloads that do not benefit much from simultaneous multithreading are those in which the majority of individual software threads use a large amount of any resource in the processor or memory. For example, workloads that are floating-point intensive are likely to gain little from simultaneous multithreading and are the ones most likely to lose performance.

IBM has documented a 25% to 40% increase in commercial workloads throughput using simultaneous multithreading.

For more information, see the following URL:

http://www.ibm.com/servers/eserver/pseries/hardware/system_perf.html

2.5 POWER Hypervisor

Combined with features designed into the POWER5 processor, the POWER Hypervisor delivers functions that enable other system technologies, including Micro-Partitioning, virtualized processors, IEEE VLAN-compatible virtual switch, virtual SCSI adapters, and virtual consoles. The POWER Hypervisor is a component of system firmware that is always active, regardless of the system configuration.

The POWER Hypervisor performs the following tasks:

- Provides an abstraction layer between the physical hardware resources and the logical partitions using them.
- Enforces partition integrity by providing a security layer between logical partitions.
- Controls the dispatch of virtual processors to physical processors.
- Saves and restores all processor state information during logical processor context switch.
- Controls hardware I/O interrupt management facilities for logical partitions.
- Provides virtual LAN channels between physical partitions that help to reduce the need for physical Ethernet adapters for interpartition communication.

Three types of virtual I/O adapters are supported by the POWER Hypervisor, which we discuss in the following sections.

2.5.1 Virtual SCSI

The POWER5 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.

2.5.2 Virtual Ethernet

The POWER Hypervisor provides a virtual Ethernet switch function that allows partitions on the *same server* a means for fast and secure communication. Virtual Ethernet working on LAN technology allows a transmission speed in the range of 1 to 3 GBps depending on the MTU[4] size. Virtual Ethernet requires a POWER5 system with either AIX 5L Version 5.3 or the appropriate level of Linux and an HMC to define the virtual Ethernet devices. Virtual Ethernet does not require the purchase of any additional features or software such as the Advanced POWER Virtualization feature.

Virtual Ethernet features include:

- A partition supports 256 virtual Ethernet connections, where a single virtual Ethernet resource can be connected to another virtual Ethernet, a real network adapter, or both in a partition. Each virtual Ethernet adapter can also be configured as a trunk adapter.

[4] Maximum transmission unit

- Each partition operating system detects the virtual local area network (VLAN) switch as an Ethernet adapter, without the physical link properties and asynchronous data transmit operations. Layer-2 bridging to a physical Ethernet adapter is also included in the virtual Ethernet features. The virtual Ethernet network is extendable outside the server to a physical Ethernet network.

> **Note:** Virtual Ethernet 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.

2.5.3 Virtual (TTY) console

Each partition needs to have access to a system console. Tasks such as operating system installation, 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 TTY, or VTerm, does not require the purchase of any additional features or software such as the Advanced POWER Virtualization feature.

Depending on the system configuration, the operating system console can be provided by the Hardware Management Console virtual TTY, or from a terminal emulator connected to an synchronous adapter in an I/O drawer.

> **Note:** The POWER5 Hypervisor is active when the server is running in partition and non-partition mode, and also when not connected to the HMC. Consider the Hypervisor memory requirements when planning the amount of system memory required. For planning information, see:
>
> `http://www.ibm.com/servers/eserver/iseries/lpar/systemdesign.htm`
>
> In AIX 5L Version 5.3, the **`lparstat`** command using the **`-h`** and **`-H`** flags displays Hypervisor statistical data. Using the **`-h`** flag adds summary Hypervisor statistics to the default **`lparstat`** output. For more information about this command, go to “The lparstat command” on page 144.

2.6 Workload and partition management

AIX 5L offers two methods of vertical server consolidation: workload management and partitioning, of which the most recent development is shared processor logical partitions (Micro-Partitioning technology) with Partition Load Manager.

This section gives you information about the software related to this topic:

- Partition Load Manager
- IBM Tivoli® Provisioning Manager
- AIX Workload Manager
- Enterprise Workload Manager

2.6.1 Partition Load Manager

The Partition Load Manager is a resource manager that assigns and moves resources based on defined policies and utilization of the resources. It helps clients to maximize the utilization of processor and memory resources of DLPAR-capable logical partitions running AIX 5L on pSeries servers.

PLM manages memory, both dedicated and shared processors, and partitions using Micro-Partitioning technology to readjust the resources. This adds additional flexibility on top of the micro-partitions flexibility added by the POWER Hypervisor.

Other key Partition Load Manager features include:

- Increased CEC resource utilization
- Demand-based and automated resource management
- Support for either dedicated or shared processor LPARs

There are some restrictions for this product:

- Response time goals are not available.
- Cannot control the addition or removal of LPAR resources.

For more information about Partition Load Manager (PLM), go to 6.5, “Partition Load Manager” on page 211.

2.6.2 IBM Tivoli Provisioning Manager

Provisioning is the process of providing IT resources to enable business functions to run. The provisioning goal is to be able to have an application running as quickly as possible with as little manual intervention as possible. Systems provisioning, powered by Tivoli Provisioning Manager, provides for rapid provisioning across an environment and it is the most comprehensive solution available from IBM.

It supports the separation of physical resources so that the IT infrastructure can be based on a pool of virtual resources. Tivoli Provisioning Manager uses predefined procedures known as workflows to automatically create, install, and configure partitions. On pSeries, these partitions can run AIX 5L or Linux. Tivoli Provisioning Manager uses the tools provided by pSeries (the HMC and NIM) to perform these tasks.

Other key Tivoli Provisioning Manager features include:

- Allows resources to be dynamically assigned to workloads that need capacity.
- Allows for sharing of resources between different workloads.

2.6.3 AIX Workload Manager

The AIX 5L Workload Manager provides system administrators a task and policy-based method for managing the resources used by applications and users within one AIX image. Workload Manager delivers automated resource administration for multiple applications running on a single server. This capability helps to ensure that critical applications are not impacted by the resource requirements of less critical jobs. Workload Manager helps deliver the benefits of server consolidation and centralized systems administration.

Workload Manager is dynamic LPAR-aware and makes the appropriate adjustments when LPAR resources are added or removed; however, it currently does not control the addition or removal of LPAR resources.

Other key Workload Manager features include:

- Support for use of either Web-based SM graphical interface or SMIT menus to create profiles (job classes) and manage job assignment
- Creation of profiles to allocate processor, memory, and disk I/O resource usage with minimum and maximum limits
- Assignment of jobs to profiles based on user/group ownership, application name or tag, process type, or manual assignment
- Automatic event notification for system administrators with ability to define responsive actions
- Flexible policy definition for processors associated with a Workload Manager class
- Prioritization of applications

There are some restrictions for this product:

- Response time goals are not available.
- Cannot control the addition or removal of LPAR resources.
- Workload management only within one AIX 5L image.

2.6.4 Enterprise Workload Manager

Enterprise Workload Manager monitors workload across a set of heterogeneous environments based on user-defined business goals and offers a single view of workload activity.

Other key Enterprise Workload Manager features include:

- Goal-based resource optimization
- End-to-end topology view and statistics for business transactions
- Improved effectiveness of physical resources
- Workload management of heterogeneous environments
- End-to-end response time reporting

2.7 Guidelines for dynamic LPAR and virtualization

This section provides information about dynamic LPAR and virtualization.

Table 2-15 lists AIX 5L and Linux support for dynamic LPAR and virtualization.

Table 2-15 Operating system supported function

Function	AIX 5L Version 5.2	AIX 5L Version 5.3	Linux SLES 9	Linux RHEL AS 3	Linux RHEL AS 4
Dynamic LPAR					
Processor	Y	Y	Y	N	Y
Memory	Y	Y	N	N	N
I/O	Y	Y	Y	N	Y
Virtualization					
Micro-partitions (1/10th of processor)	N	Y	Y	Y	Y
Virtual Storage	N	Y	Y	Y	Y
Virtual Ethernet	N	Y	Y	Y	Y
Partition Load Manager	Y	Y	N	N	N

Dynamic LPAR minimum requirements

The minimum following resources are needed per LPAR (not per system):

- At least one processor per partition for a dedicated processor partition, or at least 1/10th of a processor when using Micro-Partitioning technology.
- At least 128 MB of physical memory per additional partition.
- At least one disk (either physical or virtual) to store the operating system.
- At least one disk adapter (either physical or virtual) or integrated adapter to access the disk.
- At least one Ethernet adapter (either physical or virtual) per partition to provide a network connection to the HMC, as well as general network access.

> **Note:** We recommend that you use separate adapters for the management and the public LAN to protect the access of your system's management functions.

- A partition must have an installation method, such as NIM, and a means of running diagnostics, such as network diagnostics.

Processor

Each LPAR requires at least one physical processor if virtualization is not used. Based on this, the maximum number of dynamic LPARs without virtualization is 32 for the p5-590 server and 64 for the p5-595 server. With the use of the Advanced POWER Virtualization feature, the number of partitions per processor is 10.

Memory

It is important to highlight that the IBM @server p5 servers and their associated virtualization features have adopted an even more dynamic memory allocation policy than the previous partition-capable pSeries servers.

In a partitioned environment, some of the physical memory areas are reserved by several system functions to enable partitioning in the partitioning-capable server. You can assign unused physical memory to a partition. You do not have to specify the precise address of the assigned physical memory in the partition profile, because the system selects the resources automatically.

The Hypervisor requires memory to support the logical partitions on the server. The amount of memory required by the Hypervisor varies according to several factors. Factors influencing the Hypervisor memory requirements include:

- Number of logical partitions
- Partition environments of the logical partitions

- Number of physical and virtual I/O devices used by the logical partitions
- Maximum memory values given to the logical partitions

Generally, you can estimate the amount of memory required by server firmware to be approximately 8% of the system installed memory. The actual amount required will generally be less than 8%. However, there are some server models that require an absolute minimum amount of memory for server firmware, regardless of the previously mentioned considerations.

The minimum amount of physical memory for each partition is 128 MB, but in most cases, the actual requirements and recommendations are between 256 MB and 512 MB for AIX 5L, Red Hat, and Novell SUSE LINUX. After that, you can assign further physical memory to partitions in increments of 16 MB. This is supported for partitions running AIX 5L Version 5.2 with the 5200-04 Recommended Maintenance Package, AIX 5L Version 5.3, Red Hat Enterprise Linux AS 3 (no dynamic LPAR), Red Hat Enterprise Linux AS 4, and SUSE LINUX Enterprise Server 9. There are implications about how big a partition can grow based on the amount of memory allocated initially. For partitions that are initially sized less than 256 MB, the maximum size is 16 times the initial size. For partitions initially sized 256 MB or larger, the maximum size is 64 times the initial size.

Note: For a more detailed impression of the amount of memory required by the server firmware, use the LPAR Validation Tool (LVT). Refer to 4.3, “LPAR Validation Tool” on page 133.

I/O

The I/O devices are assigned on a slot level to the LPARs, meaning an adapter (either physical or virtual) installed in a specific slot can only be assigned to one LPAR.

If an adapter has multiple devices, such as the 4-port Ethernet adapter or the Dual Ultra3 SCSI adapter, all devices are automatically assigned to one LPAR and cannot be shared.

Devices connected to an internal controller must be treated as a group. A group can only be assigned together to one LPAR and cannot be shared.

Therefore, the following integrated devices can be independently assigned to LPARs:

- I/O drawer with Integrated Ultra320 SCSI controller

  All SCSI resources in the disk bays must be assigned together to the same LPAR. There is no requirement to assign them to a particular LPAR; in fact, they can remain unassigned if the LPAR minimum requirements are obtained using devices attached to a SCSI adapter installed in the system.

- Media devices

  The p5-590 and p-595 servers can be configured with an optional rack-mounted media drawer or storage device enclosure (IBM 7212-102).

  These devices must belong to only a single LPAR at a time; therefore, all devices in the media bays will be available to only one LPAR at a time.

Virtual I/O devices are also assigned to dynamic LPARs on a slot level. Each partition is capable to handle up to 256 virtual I/O slots. Therefore, each partition can have up to:

- 256 virtual Ethernet adapters, with each virtual Ethernet capable of being associated with up to 21 VLANs
- 256 virtual SCSI adapters

Note: For more detailed planning of the virtual I/O slots and their requirements, use the LPAR Validation Tool.

Every LPAR requires disks (either physical or virtual) for the operating system.

Partitions must be assigned to the boot adapter and disk drive from the following options:

- An internal disk drive inserted in one of the 4-pack disk bays on the I/O drawer and the SCSI controller on the drawer. Each of the disk bays is connected to a separate internal SCSI controller on the drawer.
- A boot adapter inserted in one of 20 PCI-X slots in a I/O drawer connected to the system. A bootable external disk subsystem is connected to this adapter.

Therefore, for additional LPARs without using virtualization, external disk space is necessary, which can be accomplished by using external disk subsystems. The external disk space must be attached with a separate adapter for each LPAR by using SCSI or Fibre Channel adapters, depending on the subsystem.

For additional LPARs using virtualization, the required disk drives for each partition is provided by the Virtual I/O Server partition or partitions. Physical disks owned by the Virtual I/O Server partition can either be exported and assigned to a client partition whole, or can be partitioned into several logical volumes. The logical volumes can then be assigned to different partitions.

Additional disk space can be provided by using an external storage subsystem such as the IBM TotalStorage® DS4500 for LPARs using virtualization or direct attachment.

Therefore, for additional LPARs without using virtualization, an additional Ethernet adapter is necessary. As stated previously we highly recommend that you use separate Ethernet adapters for the connection to the management LAN and public LAN.

Additional partitions using virtualization can implement the required Ethernet adapters as virtual Ethernet adapters. Virtual Ethernet adapters can be used for all kind of interpartition communication. To connect the virtual Ethernet LANs to an external network, one or more Shared Ethernet Adapters (SEAs) can be used in the Virtual I/O Server partition.

2.8 Firmware

Depending on your service environment, you can download your server firmware fixes using different interfaces and methods. The p5-590 and p5-595 servers must use the HMC to install server firmware fixes. Firmware is loaded on to the server and to the Bulk Power Controller over the HMC to the frame Ethernet network.

2.8.1 Server firmware

Server firmware is the part of the Licensed Internal Code that enables hardware, such as the service processor. Check for available server firmware fixes regularly, and download and install the fixes if necessary. The server firmware binary image is a single image that includes code for the service processor, the POWER Hypervisor, and platform partition firmware. This server firmware binary image is stored in the service processor's flash memory and executed in service processor main memory.

Because there are dual service processors per CEC, both service processors are updated when firmware updates are applied and activated using the "Licensed Internal Code Updates" section of the HMC.

Firmware is available for download at:

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

2.8.2 Power subsystem firmware

Power subsystem firmware is the part of the Licensed Internal Code that enables the power subsystem hardware in the model p5-590 and p5-595 servers. You must use an HMC to update or upgrade power subsystem firmware fixes.

The Bulk Power Controller (BPC) has its own service processor. The power firmware not only has the code load for the BPC service processor itself, but it also has the code for the distributed converter assemblies (DCAs), bulk power regulators (BPRs), fans, and other more granular field replaceable units that have firmware to help manage the frame and its power and cooling controls. The BPC service processor code load also has the firmware for the cluster switches that can be installed in the frame.

In the same way that the central electronics complex (CEC) has dual service processors, the power subsystem has dual BPCs. Both are updated when firmware changes are made using the “Licensed Internal Code Updates” section of the HMC.

The BPC initialization sequence after the reboot is unique. The BPC service processor must check the code levels of all the power components it manages, including DCAs, BPRs, fans, and cluster switches, and it must load those if they are different from what is in the active flash side of the BPC. Code is cascaded to the downstream power components over universal power interface controller (UPIC) cables.

2.8.3 Platform initial program load

The main function of the p5-590 and p5-595 service processors is to initiate platform initial program load (IPL), also referred to as platform boot. The service processor has a self-initialization procedure and then initiates a sequence of initializing and configuring many components on the CEC backplane.

The service processor has various functional states, which can be queried and reported to the POWER Hypervisor. Service processor states include, but are not limited to, standby, reset, power up, power down, and runtime. As part of the IPL process, the primary service processor will check the state of the backup. The primary service processor is responsible for reporting the condition of the backup service processor to the POWER Hypervisor. The primary service processor will wait for the backup service processor to indicate that it is ready to continue with the IPL (for a finite time duration). If the backup service processor fails to initialize in a timely fashion, the primary will report the backup service processor as a non-functional device to the POWER Hypervisor and will mark it as a GARDed resource before continuing with the IPL. The backup service processor can later be integrated into the system.

Open Firmware

IBM eserver p5 and OpenPower™ servers have one instance of Open Firmware both when in the partitioned environment and when running as a Full System Partition. Open Firmware has access to all devices and data in the system. Open Firmware is started when the system goes through a power-on reset. Open Firmware, which runs in addition to the Hypervisor in a partitioned environment, runs in two modes: global and partition. Each mode of Open Firmware shares the same firmware binary that is stored in the flash memory. In a partitioned environment, partition Open Firmware runs on top of the global Open Firmware instance. The partition Open Firmware is started when a partition is activated. Each partition has its own instance of Open Firmware and has access to all the devices assigned to that partition. However, each instance of partition Open Firmware has no access to devices outside of the partition in which it runs. Partition firmware resides within the partition memory and is replaced when AIX 5L takes control. Partition firmware is needed only for the time that is necessary to load AIX 5L into the partition system memory.

The global Open Firmware environment includes the partition manager component. That component is an application in the global Open Firmware that establishes partitions and their corresponding resources (such as CPU, memory, and I/O slots), which are defined in partition profiles. The partition manager manages the operational partitioning transactions. It responds to commands from the service processor external command interface that originate in the application that is running on the HMC.

The ASMI can be accessed during boot time or using the ASMI and selecting to boot to the Open Firmware prompt.

For more information about Open Firmware, refer to *Partitioning Implementations for IBM eserver p5 Servers,* SG24-7039, at:

http://www.redbooks.ibm.com/abstracts/sg247039.html

Temporary and permanent side of the service processor

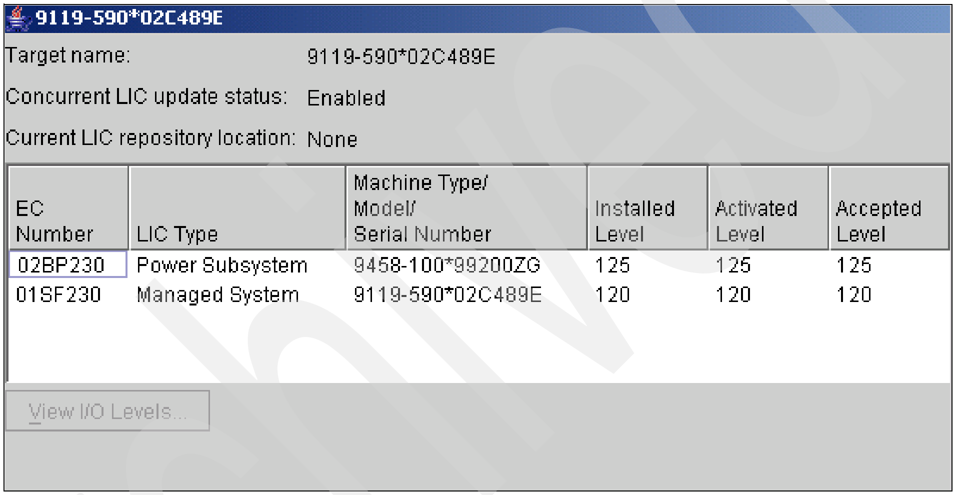
The service processor and the BPC (when present) maintain two copies of the firmware:

- One copy is considered the permanent or backup copy and is stored on the permanent side, sometimes referred to as the “p” side.
- The other copy is considered the installed or temporary copy and is stored on the temporary side, sometimes referred to as the “t” side. We recommend that you start and run the server from the temporary side.
- The copy actually booted from is called the activated level, sometimes referred to as “b”.

(The concept of “sides” is an abstraction. The firmware is located in flash memory and pointers in NVRAM determine which is “p” and “t”.)

> **Note:** The default value the system will boot is “temporary”.

To view the firmware levels on the HMC, select **Licensed Internal Code** → **Updates Change Internal Code** → **Select managed system** → **View System Information** → **None** and the window in Figure 2-6 opens. (The power subsystem is always machine type 9458, while the server is machine type 9119). We discuss this because it provides the most comprehensive example.



9119-590*02C489E

Target name: 9119-590*02C489E

Concurrent LIC update status: Enabled

Current LIC repository location: None

EC Number	LIC Type	Machine Type/ Model/ Serial Number	Installed Level	Activated Level	Accepted Level
02BP230	Power Subsystem	9458-100*99200ZG	125	125	125
01SF230	Managed System	9119-590*02C489E	120	120	120

Figure 2-6 p5-590 and p5-595 code levels

In Figure 2-6:

- The Installed Level indicates the level of firmware that has been installed and will be installed into memory after the managed system is powered off and powered on using the default “temporary” side.
- The Activated Level indicates the level of firmware that is active and running in memory.
- The Accepted Level indicates the backup level (or “permanent” side) of firmware. You can return to the backup level of firmware if you decide to remove the installed level.

The following example is the output of the **lsmcode** command for AIX 5L and Linux, showing the firmware levels as they are displayed in the outputs:

- AIX 5L:

```
The current permanent system firmware image is SF230_120
The current temporary system firmware image is SF230_120
The system is currently booted from the temporary firmware image.
```

- Linux:

```
system:SF230_120 (t) SF230_120 (p) SF230_120 (b)
```

Firmware level naming conventions

The format of the firmware level as reported by the system is:
01SF225_096 → PPNNSSS_FFF

- PP Package identifier 01 = managed system 02 = power code
- NN Machine type SF = POWER5 system BP = bulk power code
- SSS Release level
- FFF Fix pack number

Note: The following points are of special interest:

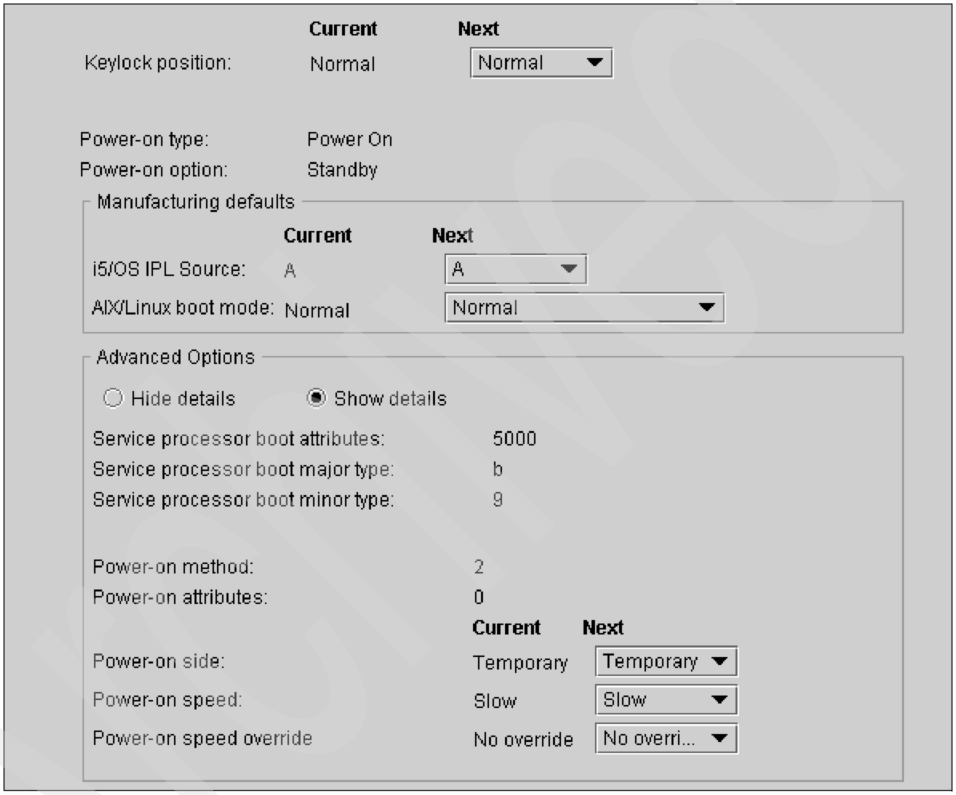
- The server firmware fix is installed on the temporary side only after the existing contents of the temporary side are permanently installed on the permanent side (the service processor performs this process automatically when you install a server firmware fix).
- If you want to preserve the contents of the permanent side, you need to remove the current level of firmware (copy the contents of the permanent side to the temporary side) before you install the fix.

However, if you get your fixes using Advanced features on the HMC interface and you indicate that you do not want the service processor to automatically accept the firmware level, the contents of the temporary side are not automatically installed on the permanent side. In this situation, you do not need to remove the current level of firmware to preserve the contents of the permanent side before you install the fix.

You might want to use the new level of firmware for a period of time to verify that it works correctly. When you are sure that the new level of firmware works correctly, you can permanently install the server firmware fix. When you permanently install a server firmware fix, you copy the temporary firmware level from the temporary side to the permanent side.

Conversely, if you decide that you do not want to keep the new level of server firmware, you can remove the current level of firmware. When you remove the current level of firmware, you copy the firmware level that is currently installed on the permanent side is copied from the permanent side to the temporary side.

Choosing which firmware to use when powering on the system is done using the Power-On Parameters tab in the server properties box, as shown in Figure 2-7 on page 29.



	Current	Next
Keylock position:	Normal	Normal ▼

Power-on type: Power On
Power-on option: Standby

Manufacturing defaults

	Current	Next
i5/OS IPL Source:	A	A ▼
AIX/Linux boot mode:	Normal	Normal ▼

Advanced Options

○ Hide details ● Show details

Service processor boot attributes: 5000
Service processor boot major type: b
Service processor boot minor type: 9

Power-on method: 2
Power-on attributes: 0

	Current	Next
Power-on side:	Temporary	Temporary ▼
Power-on speed:	Slow	Slow ▼
Power-on speed override	No override	No overri... ▼

Figure 2-7 Power-on parameters

Get server firmware fixes using an HMC

Periodically, you need to download and install fixes for your server and power subsystem firmware.

How you get the fix depends on whether or not the HMC or server is connected to the Internet:

- If the HMC or server is connected to the Internet:

  There are several repository locations from which you can download the fixes using the HMC. For example, you can download the fixes from your service provider's Web site or support system, from optical media that you order from your service provider, or from an FTP server on which you previously placed the fixes.

- If neither the HMC nor your server is connected to the Internet:

  You need to download your new system firmware level to a CD-ROM media or FTP server.

For both of these options, you can use the interface on the HMC to install the firmware fix (from one of the repository locations or from the optical media). The Change Internal Code wizard on the HMC provides a step-by-step process for you to perform the required steps to install the fix:

1. Ensure that you have a connection to the service provider (if you have an Internet connection from the HMC or server).
2. Determine the available levels of server and power subsystem firmware.
3. Create the optical media (if you do not have an Internet connection from the HMC or server).
4. Use the Change Internal Code wizard to update your server and power subsystem firmware.

   > **Note:** The tasks in the “Licensed Internal Code Updates” view on the HMC vary according to the HMC code level:
   >
   > - Version 4.4.x and earlier:
   >   - Use “Change Internal Code” to update within a release.
   >   - Use “Manufacturing Equipment Specification Upgrade” to upgrade to a new release.
   > - Version 4.5.x and later:
   >   - Use “Change Licensed Internal Code for the current release” to update within a release.
   >   - Use “Upgrade Licensed Internal Code to a new release” to upgrade to a new release.

5. Verify that the fix installed successfully.

> **Note:** To view existing levels of server firmware using the `lsmcode` command, you need to have the following service tools installed on your server:
>
> - AIX 5L
>
>   You must have AIX 5L diagnostics installed on your server to perform this task. AIX 5L diagnostics are installed when you install the AIX 5L operating system on your server. However, it is possible to deselect the diagnostics. Therefore, you need to ensure that the online AIX 5L diagnostics are installed before proceeding with this task.
>
> - Linux
>
>   - Platform Enablement Library: librtas-xxxxx.rpm
>   - Service Aids: ppc64-utils-xxxxx.rpm
>   - Hardware Inventory: lsvpd-xxxxx.rpm
>
>   Where xxxxx represents a specific version of the RPM file.
>
>   If you do not have the service tools on your server, you can download them from the following Web page:
>
>   `http://techsupport.services.ibm.com/server/lopdiags`

2.9 Capacity on Demand

Through unique CoD offerings, IBM servers can offer either permanent or temporary increases in processor and memory capacity. CoD is available in four activation configurations, each with specific pricing and availability terms. In this section, we discuss the four types of CoD activation configurations from a functional standpoint. Contractual and pricing issues are outside the scope of this document and should be discussed with either your IBM Global Financing Representative, IBM Business Partner, or IBM Sales Representative.

Capacity on Demand is supported by the following operating systems:

- AIX 5L Version 5.2 and Version 5.3
- i5/OS V5R3, or later
- SUSE LINUX Enterprise Server 9 for POWER, or later
- Red Hat Enterprise Linux AS 3 for POWER (update 4), or later

For additional information about Capacity on Demand, see:

`http://www.ibm.com/servers/eserver/pseries/ondemand/cod/`

2.9.1 Types of Capacity on Demand

Capacity on Demand for the @server p5 systems with dynamic logical partitioning (dynamic LPAR) offers system owners the ability to non-disruptively activate processors and memory without rebooting partitions. CoD also gives @server p5 systems owners the option to temporarily activate processors to meet varying performance needs and to activate additional capacity on a trial basis.

IBM has established four types of CoD offerings, each with a specific activation plan. Providing different types of CoD offerings gives clients flexibility when determining their resource needs and establishing their IT budgets. IBM Global Financing can help match individual payments with capacity usage and competitive financing for fixed and variable costs related to Capacity on Demand offerings. By financing Capacity on Demand costs and associated charges together with a base lease, spikes in demand need not become spikes in a budget.

After a system with CoD features is delivered, it can be activated in the following ways:

- Capacity Upgrade on Demand (CUoD) for processors and memory
- On/Off Capacity on Demand (CoD) for processors and memory
- Reserve Capacity on Demand (CoD) for processors only
- Trial Capacity on Demand (CoD) for processors and memory

The servers use specific feature codes to enable CoD capabilities. All types of CoD transactions for processors are in whole numbers of processors, not in fractions of processors. All types of CoD transactions for memory are in 1 GB increments.

Table 2-16 on page 60 provides a brief description of the four types of CoD offerings, identifies the proper name of the associated activation plan, and indicates the default type of payment offering and scope of enablement resources. The payment offering information is intended for reference only; all pricing agreements and service contracts should be handled by your IBM representative. The subsequent sections provide a functional overview of each CoD offering.

Table 2-16 Types of Capacity on Demand (functional categories)

Activation plan	Functional category	Applicable system resources	Type of payment offering	Description
Capacity Upgrade on Demand	Permanent capacity for nondisruptive growth	Processor and memory resources	Pay when purchased	Provides a means of planned growth for clients who know they will need increased capacity but are not sure when.
On/Off Capacity on Demand	Temporary capacity for fluctuating workloads	Processor and memory resources	Pay after activation	Provides for planned and unplanned short-term growth driven by temporary processing requirements such as seasonal activity, period-end requirements, or special promotions.
Reserve Capacity on Demand		Processor resources only	Pay before activation	
Trial Capacity on Demand	Temporary capacity for workload testing or any one time need	Processor and memory resources	One-time, no-cost activation for a maximum period of 30 consecutive days	Provides the flexibility to evaluate how additional resources will affect existing workloads, or to test new applications by activating additional processing power or memory capacity (up to the limit installed on the server) for up to 30 contiguous days.
Capacity Backup	Disaster recovery	Off-site machine	Pay when purchased	Provides a means to purchase a machine for use when off-site computing is required, such as during disaster recovery.

Capacity Upgrade on Demand (CUoD) for processors

Capacity Upgrade on Demand (CUoD) for processors allows inactive processors to be installed in the server and can be permanently activated by the client as required.

All processor books available on the p5-590 and p5-595 are initially implemented as 16-way CoD offerings with zero active processors.

A minimum of 8 or 16 permanently activated processors are required on the p5-590 or p5-595 server.

The number of permanently activated processors is based on the number of processor books installed:

- One processor book installed requires 8 (p5-590) or 16 (p5-595) permanently activated processors.
- Two processor books installed requires 16 permanently activated processors.
- Three processor books installed requires 24 permanently activated processors.
- Four processor books installed requires 32 permanently activated processors.

Additional processors on the CoD books are activated in increments of one by ordering the appropriate activation feature number. If more than one processor is to be activated at the same time, the activation feature should be ordered in multiples.

After receiving an order for a CUoD for the processor activation feature, IBM will provide the client with a 34-character encrypted key. This key is entered into the system to activate the desired number of additional processors.

CUoD processors that have not been activated are available to the p5-595 server for dynamic processor sparing when running the AIX 5L operating system. If the server detects the impending failure of an active processor, it will attempt to activate one of the unused CoD processors and add it to the system configuration. This helps to keep the server's processing power at full strength until a repair action can be scheduled.

Capacity Upgrade on Demand for memory

Capacity Upgrade on Demand (CUoD) allows inactive memory to be installed in the p5-590 or p5-595 server and can be permanently activated by the client as required.

CUoD for memory can be used in any available memory position.

Additional CoD memory cards are activated in increments of 1 GB by ordering the appropriate activation feature number. If more than one 1 GB memory increment is to be activated at the same time, the activation code should be ordered in multiples.

After receiving an order for a CUoD for memory activation feature, IBM will provide the client a 34-character encrypted key. This key is entered into the system to activate the desired number of additional 1 GB memory increments.

Memory configuration rules for the p5-590 and p5-595 servers apply to CUoD for memory cards as well as conventional memory cards. The memory configuration rules are applied based on the maximum capacity of the memory card:

- Apply 4 GB configuration rules for 4 GB CoD for memory cards with less than 4 GB of active memory.
- Apply 8 GB configuration rules for 8 GB CoD for memory cards with less than 8 GB of active memory.

On/Off Capacity on Demand (On/Off CoD)

On/Off Capacity on Demand (On/Off CoD) allows clients to temporarily activate installed CUoD processors and memory resources and later deactivate the resources as desired.

On/Off processor and memory resources is implemented on a *pay-as-you-go* basis using:

- On/Off Processor and Memory Enablement features.

  Signing an On/Off Capacity on Demand contract is required. An enablement code will be supplied to activate the enablement feature.

- After the On/Off Enablement feature is ordered and the associated enablement code is entered into the system, the client must report on/off usage to IBM at least monthly. This information, which is used to compute the billing data on a quarterly basis, is provided to the sales channel, which will place an order for the quantity of On/Off Processor Day and Memory Day billing features used and invoice the client.

Each On/Off CoD activation feature provides 30 days of usage for the CUoD processor or memory feature. Each On/Off CoD feature applies to any of the inactive CUoD processors or memory installed in the system. The following examples are of the usage calculation:

- A processor day is measured each time a processor is activated in a 24-hour period or any fraction hours of testing or production. The result is four processor usage days.
- A new measurement day starts each time processors are activated. If a client activates four processors for a two-hour test and later in the same 24-hour period activates two processors for two hours to meet a peak workload, the result is six processor days usage.

Reserve Capacity on Demand (Reserve CoD)

Reserve Capacity on Demand (Reserve CoD) is available for p5-590 and p5-595 servers. Reserve CoD is an innovative offering allowing clients to temporarily activate in an automated manner installed CoD processors used within a shared processor pool. Charges for the temporary activation of Reserve CoD processors are only incurred when processing needs exceed the fully entitled level.

Reserve CoD is a pre-pay method of temporary activation. It is ordered by purchasing the quantity of Reserve CoD features appropriate for the model and speed of installed processors. Each feature includes 30 days of temporary usage time. When Reserve CoD is ordered, the user will receive a 34-digit activation code to be entered at the HMC. The activation code establishes the Reserve CoD balance of available usage time. Inactive CoD processors can then be assigned to the shared processor pool, which will be available for workload processing. Charges for the inactive processors will only be incurred when the workload in the shared pool exceeds 100% of the entitled (permanently activated) level of performance. Charges are made against the Reserve CoD account balance in increments of processor days, and Advanced POWER Virtualization must be activated in order to use Reserve CoD.

Trial Capacity on Demand

Trial Capacity on Demand (Trial CoD) is a function delivered with all pSeries servers supporting CUoD resources beginning May 30, 2003. Those servers with standby CoD processors or memory will be capable of using a one-time, no-cost activation for a maximum period of 30 consecutive days. This enhancement allows for benchmarking of CoD resources or can be used to provide immediate access to standby resources when the purchase of a permanent activation is pending.

Trial CoD is a complimentary service offered by IBM. Although IBM intends to continue it for the foreseeable future, IBM reserves the right to withdraw Trial CoD at any time, with or without notice.

Capacity BackUp

Also available are three new Capacity BackUp features for configuring systems used for disaster recovery. Capacity BackUp for IBM eserver p5 590 and 595 systems offers an off-site, disaster recovery machine at an affordable price. This disaster recovery machine has primarily inactive Capacity on Demand (CoD) processors that can be activated in the event of a disaster. Capacity BackUp for IBM eserver p5 offering includes:

- Four processors that are permanently activated and can be used for any workload
- Either 28 or 60 standby processors to be used in the event of a disaster
- Either 900 (4/32-way) or 1800 (4/64-way) of On/Off CoD processor days available for testing or for use in the event of a disaster

Capacity BackUp systems can be turned on at any time by using the On/Off CoD activation procedure for the needed performance during an unplanned system outage. Each Capacity BackUp configuration is limited to 450 On/Off CoD credit days per processor book. For clients who require additional capacity or processor days, additional processor capacity can be purchased under IBM CoD at regular On/Off CoD activation prices. IBM HACMP V5 and HACMP/XD software (5765-F62), when installed, can automatically activate Capacity BackUp resources on failover. When needed, HACMP can also activate dynamic LPAR and CoD resources.

3

System partitioning and operating system installation

This chapter describes the installation and configuration of logical partitions. Partitioning enables users to configure a single computer into several independent systems. Each partition is capable of running applications in its own independent environment. Unless the system is client of a Virtual I/O Server, it contains its own operating system, set of system processors, set of system memory, and I/O adapters. If it is a Virtual I/O Server client, network adapters, as well as the physical disk where the operating system is installed, might be virtual devices from a Virtual I/O Server. Such systems will still behave as though they use dedicated devices.

This chapter contains the following sections:

- Initial system configuration
- Basic system management tasks
- Logical partitioning a system
- LPAR resource administration
- Installation of the operating system into a partition
- Dynamic LPAR operation

3.1 Initial system configuration

There are things to consider when planning the initial installation of your IBM eServer p5 server. Use this section as a guide to setup and customize your system. This section contains information about the Advanced System Management Interface (ASMI), system management services, and basic system management tasks.

Important: It is very important that you do not attempt to power up a new enterprise server without a correctly installed HMC.

3.1.1 Advanced System Management Interface

The Advanced System Management Interface (ASMI) is an interface to the service processor that enables you to set flags that affect the operation of the server, such as auto power restart, and to view information about the server, such as the error log and vital product data.

This interface is accessible using a Web browser on a client system that is connected to the service processor on an Ethernet network. It can also be accessed using the HMC. The service processor and the ASMI are standard on all IBM eServer i5, eServer p5, and OpenPower servers.

Accessing the ASMI using a Web browser

The Web interface to the Advanced System Management Interface is accessible through Microsoft Internet Explorer 6.0, Netscape 7.1, or Opera 7.23 running on a PC or mobile computer connected to the service processor. The Web interface is available during all phases of system operation including the initial program load and run time. However, some of the menu options in the Web interface are unavailable during IPL or run time to prevent usage or ownership conflicts if the system resources are in use during that phase.

The following instructions apply to p5-590/p5-595 systems that are disconnected from a Hardware Management Console (in most cases because of an error), or if you need to change the network configuration in the service processor and the HMC is unavailable. This is normally a serious condition left to trained service engineers. These instructions also apply to other eServer p5 servers.

If you are managing the server using an HMC, access the ASMI using the HMC. Complete the following tasks to set up the Web browser for direct or remote access to the ASMI:

1. Connect the power cord from the server to a power source, and wait for the STBY LED (on the Bulk Power Controller) to be on solid.

2. Select a PC or mobile computer that has Microsoft Internet Explorer 6.0, Netscape 7.1, or Opera 7.23 to connect to your server. If you do not plan to connect your server to your network, this PC or mobile computer will be your ASMI console. If you plan to connect your server to your network, this PC or mobile computer will be temporarily connected directly to the server for setup purposes only. After setup, you can use any PC or mobile computer on your network that is running Microsoft Internet Explorer 6.0, Netscape 7.1, or Opera 7.23 as your ASMI console.
3. Determine the IP address of the Ethernet port to which your PC or mobile computer is connected, and enter the IP address in the address field of the Web browser, for example, 192.168.2.130. You can use a standard or a crossover Ethernet cable.

   For example, if you connected your PC or mobile computer to Port A, enter the following IP address in your Web browser address field:

   ```
   https://192.168.2.147
   ```

   For all other eserver systems, connect you PC or mobile computer to the HMC2 port on the service processor. The standard address is:

   ```
   https://192.168.3.147
   ```

Accessing ASMI using HMC or Web-based System Manager

Note: Since HMC Machine Code V4.5, the ASMI menu works with the Web-based System Manager as well.

To access the Advanced System Management Interface using the Hardware Management Console, complete the following steps:

1. Ensure that the HMC is set up and configured.
2. In the Navigation Area, expand the managed system with which you want to work.
3. Expand **Service Applications** and click **Service Focal Point**.
4. In the content area, click **Service Utilities**.
5. From the Service Utilities window, select the managed system with which you want to work.
6. From the Selected menu on the Service Utilities window, select **Launch ASMI**.

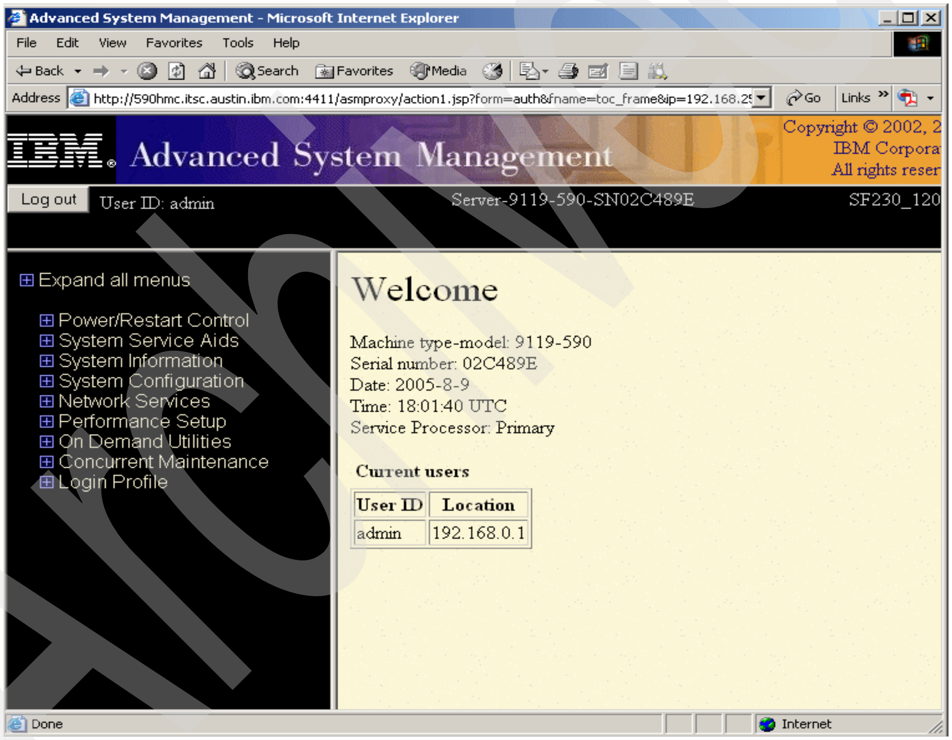
ASMI user accounts

ASMI enables you to create several types of users. In Table 3-1 provides the default user login and passwords. Usually, you will use the admin login; however, be aware that the default passwords can be changed.

Table 3-1 ASMI user accounts

User ID	Password
admin	admin
general	general

Figure 3-1 shows the ASMI menu with all the functions.



http://590hmc.itsc.austin.ibm.com:4411/asmproxy/action1.jsp?form=auth&fname=toc_frame&ip=192.168.2!

User ID: admin

Server-9119-590-SN02C489E

SF230_120

Welcome

Machine type-model: 9119-590
Serial number: 02C489E
Date: 2005-8-9
Time: 18:01:40 UTC
Service Processor: Primary

Current users

User ID	Location
admin	192.168.0.1

Figure 3-1 ASMI menu

3.1.2 System Management Services

System tasks and monitoring tasks can be accomplished using the System Management Service (SMS).

To start the system in SMS mode using the HMC, perform the following steps:

1. Use the HMC to activate the server or partition and select **Advanced**.
2. From the Boot Mode pull-down menu, select **SMS**.

Using System Management Services

The System management Services (SMS) menus can be used to view information about your system or partition and to perform tasks such as setting a password, changing the boot list, and setting the network parameters.

> **Note:** In a partitioned system, only the devices assigned to the partition being booted display in the SMS menus. In a Full System Partition, all devices in the system display in the SMS menus.

To start the system management services using a terminal or vterm, perform the following steps:

1. For a partitioned system, use the Hardware Management Console to restart the partition.
2. For a partitioned system, watch the virtual terminal window on the HMC.
3. Look for the POST indicators, `memory`, `keyboard`, `network`, `SCSI`, `speaker`, which appear across the bottom of the screen.

   Press the numeric 1 key after the word `keyboard` appears and before the word `speaker` appears.

After the system management services starts, a screen similar to the one shown in Example 3-1 opens.

Example 3-1 SMS main menu

```
                                   Main Menu

 1   Select Language
 2   Change Password (Options NOT available in LPAR mode)
 3   View Error Log (Option NOT available in LPAR mode)
 4   Setup Remote IPL (Initial Program Load)
 5   Change SCSI Settings
 6   Select Console
 7   Select Boot Options
-----------------------------------------------------------------------------
Navigation keys:
                                         X = eXit System Management Services
-----------------------------------------------------------------------------
Type the number of the menu item and press Enter or select a Navigation key: _
```

We describe each option on the system management main menu in this section.

Select Language

This option enables you to change the language used by the text-based System Management Services menus.

Change Password Options

The Change Password Options menu enables you to select from the following password utilities:

- Set Privileged-Access Password
- Remove Privileged-Access Password

The privileged-access password protects against the unauthorized start of the system programs. If the privileged-access password has been enabled, the system will prompt for the password whenever it reboots.

View Error Log

When you select this option, you can view or clear your system error messages.

Setup Remote IPL (Initial Program Load)

This option allows you to enable and set up the remote startup of your system or partition. You must first specify the network parameters. This is typically used to do a ping test to see if a Network Installation Management (NIM) client can communicate with a NIM server.

Change SCSI Settings

SCSI utilities enable you to set delay times for the SCSI hard disk spin-up and to set SCSI IDs for SCSI controllers installed in the system.

Select Console

Select this option to define which display is used by the system for system management. If no console is selected, the console defaults to serial port 1 on the primary I/O book.

> **Note:** This option is not available on partitioned systems. A virtual terminal window on the HMC is the default firmware console for a partitioned system.

Select Boot Options

Use this menu to view and set various options regarding the installation and boot devices:

- Select Install or Boot a Device: Enables you to select a device from which to boot or install the operating system. This option is for the current boot only.
- Select Boot Devices: Select this option to view and change the customized boot list, which is the sequence of devices that the system searches when booting an operating system. The boot device list can contain up to five devices.
- Multiboot Startup: Toggles the multiboot startup flag, which controls whether the multiboot menu is invoked automatically on startup.

> **Note:** In a partitioned system, only those devices from which an operating system can be booted that are assigned to the partition that is being booted display on the Select Boot Devices menu. In a Full System Partition, devices from which an operating system can be booted display on the Select Boot Devices menu.

Exiting System Management Services

After you have finished using the system management services, type `x` (for exit) to boot your system or partition.

3.2 Basic system management tasks (HMC)

This section describes how to perform the following system operations:

- Managing the system and a frame
- Powering on the managed system

3.2.1 Managing the system

You can perform the tasks discussed in this chapter when the managed system is selected in the Contents area. The managed system (actually the CEC) is shown below a frame in the Contents area.

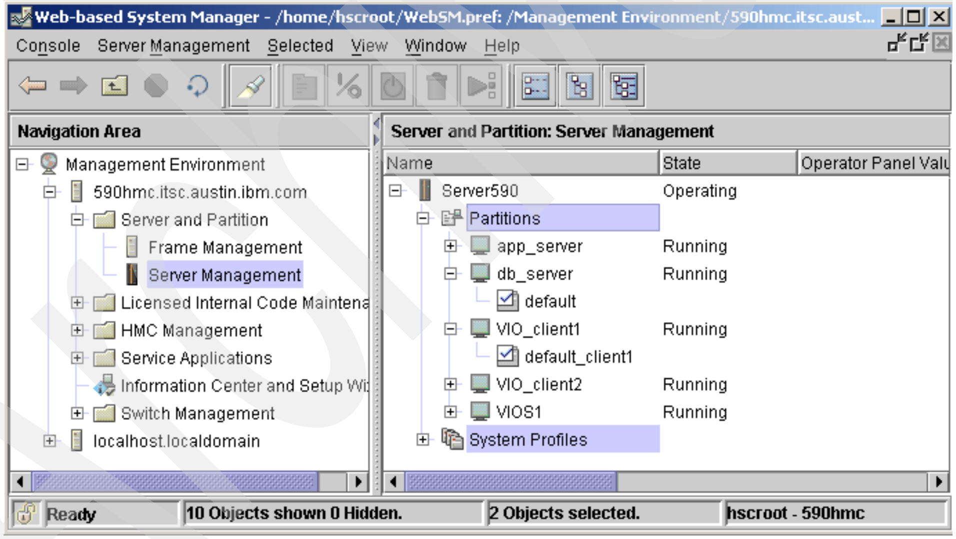
The HMC communicates with the managed system to perform various system management service and partitioning functions. Systems connected to an HMC are recognized automatically by the HMC and are shown in the Contents area.

You can connect up to two HMC in the following ways on a p5-590 or p5-595:

- The first (the main or only) HMC is connected using a private network to BPC-A (Bulk Power Controller). The HMC must be set up to provide DHCP addresses on that private (eth0) network.
- A secondary (redundant) HMC is connected using a separate private network to BPC-B. That second HMC must be set up to use a different range of addresses for DHCP.
- Additional provisions must be made for an HMC connection to the BPC in a powered expansion frame.

On other systems, the HMC is connected to one of the available HMC ports.

To view more information about the managed system, select the **Server Management** icon in the Navigation Area. Figure 3-2 shows an example of server manager. The Contents area expands to show a frame, which you can then expand to show information about the managed system, including its name, its state, and the operator panel value.


Web-based System Manager - /home/hscroot/WebSM.pref: /Management Environment/590hmc.itsc.aust...
Console
Server Management
Selected
View
Window
Help
Navigation Area
Management Environment
590hmc.itsc.austin.ibm.com
Server and Partition
Frame Management
Server Management
Licensed Internal Code Maintena
HMC Management
Service Applications
Information Center and Setup Wi
Switch Management
localhost.localdomain
Server and Partition: Server Management

Name	State	Operator Panel Valu
Server590	Operating	
Partitions		
app_server	Running	
db_server	Running	
default		
VIO_client1	Running	
default_client1		
VIO_client2	Running	
VIOS1	Running	
System Profiles		

Ready | 10 Objects shown 0 Hidden. | 2 Objects selected. | hscroot - 590hmc

Figure 3-2 The Server Management window

To expand the view of the managed system's properties, click the plus sign (**+**) next to the managed system's name to view its contents.

In the Contents area, you can also select the managed system by right-clicking the managed system icon to perform the following tasks:

- Power the managed system on or off.

- View the managed system's properties.
- Open and close a terminal window.
- Create, restore, back up, and remove system profile data.
- Manage on demand activations.
- Rebuild the managed system.
- Release the HMC lock on this managed system.
- Delete the managed system from the HMC graphical user interface.

You can also access these options by clicking the managed system and clicking **Selected** from the menu.

Powering on the managed system

You can use your HMC to power on the managed system.

You must be a member of one of the following roles: System Administrator, Advanced Operator, Service Representative, or Operator.

To power on the managed system, perform the following steps:

1. In the Navigation Area, click the **Partition Management** icon.
2. In the Contents area, select the managed system.
3. From the menu, select **Selected**.
4. Select **Power On**.

A panel opens that offers the four power-on modes: System Profile, Full System Partition, Partition Standby, and Auto Start Partitions.

Note: You must power off your managed system to switch between using the Full System Partition and using logical partitions.

System Profile

The System Profile option powers on the system according to a predefined set of profiles. The profiles are activated in the order in which they are shown in the system profile.

Partition Standby

The Partition Standby power-on option enables you to create and activate logical partitions. When the Partition Standby power-on is completed, the operator panel on the managed system displays `LPAR...`, indicating that the managed system is ready for you to use the HMC to partition its resources.

> **Note:** The Full System Partition is displayed as Not Available because the managed system was powered on using the Partition Standby option.

Auto Start Partitions

The option powers on the managed system to partition standby mode and activates all partitions that have been powered on by the HMC at least once. For example, if you create a partition with four processors, use the DLPAR operation to remove one processor, and then shut down the system, the Auto Start Partitions power-on mode activates this partition with three processors. This is because the 3-processor configuration was the last configuration used, and the HMC ignores whatever you have specified in the partition's profile. Using this option, the activated partitions boots the operating system using a normal mode boot, even if the default profile for the partition specifies the other modes, such as boot to SMS.

> **Note:** This power-on mode is available on the HMC software release beginning with Release 3, Version 2.

Power-on options

The Full System Partition has predefined profiles. You cannot change, add, or delete them. The predefined profiles are as follows:

- Power On Normal

  This boots an operating system from the designated boot device.

- Power On Diagnostic Stored Boot List

  This causes the system to perform a service mode boot using the service mode boot list saved on the managed system. If the system boots AIX 5L from the disk drive, and AIX 5L diagnostics are loaded on the disk drive, AIX 5L boots to the diagnostics menu.

  Using this option to boot the system is the preferred way to run online diagnostics.

- Power On SMS

  This boots to the System Management Services (SMS) menus.

- Power On Diagnostic Default Boot List

  Similar to Power On Diagnostic Stored Boot List Profile, except the system boots using the default boot list that is stored in the system firmware. This is normally used to try to boot diagnostics from the CD-ROM drive.

  Using this option to boot the system is the preferred way to run stand-alone diagnostics.

- Power On Open Firmware OK Prompt

  When this selection is enabled, the system boots to the Open Firmware prompt.

Powering off the managed system

You can also use your HMC to power off the managed system. Ensure that all partitions have been shut down and their states have changed from Running to Ready.

To power off the managed system, you must be a member of one of the following roles: System Administrator, Advanced Operator, or Service Representative.

To power off the managed system, perform the following steps:

1. In the Contents area, select the managed system.
2. From the menu, select **Selected**.
3. Select **Power Off**.

If you attempt to power off a system that has active partitions, you will receive a warning to that effect, but you will still be able to power off the managed system. Each partition associated with that managed system will also power off. That might be a fatal exercise to applications running on the partition.

Viewing managed system properties

To view your managed system's configuration and capabilities, use the Properties window.

Any user can view managed system properties.

To view your managed system's properties, perform the following steps:

1. In the Contents area, select the managed system.
2. From the menu, select **Selected**.
3. Select **Properties**.

If you have powered on your system using the Full System Partition option, the HMC displays the system's name, partition capability, state, serial number, model and type, and policy information. A system that is powered on using the Partition Standby option displays this information, as well as available and assigned processors, memory, I/O drawers and slots, and policy information. The Processor tab displays information that is helpful when performing dynamic logical partitioning processor tasks.

Use the Processor tab to view the processor status, the processor state, and whether a processor is assigned to a partition. The information in the Processor tab is also helpful when you need to know if processors are disabled and therefore are not able to be used by any partition.

Notes:

- You can back up, restore, initialize, and remove profiles that you have created. Also, you can release an HMC lock on the managed system if you have two HMCs connected to your managed system and one of the HMCs is not responding.
- You can access most menu options by right-clicking.

3.2.2 Rebuilding the managed system

Rebuilding the managed system acts much like a refresh of the managed system information. Rebuilding the managed system is useful when the system's state indicator in the Contents area is shown as Recovery. The Recovery indicator signifies that the partition and profile data stored in the managed system must be refreshed. It might also be required when major components of the system, such as I/O drawers, are replaced for service reasons.

This operation is different from performing a refresh of the local HMC panel. In this operation, the HMC reloads information stored on the managed system.

Any user can rebuild the managed system.

To rebuild the managed system, perform the following steps:

1. In the Contents area, select the managed system.
2. From the menu, select **Selected**.
3. Select **Rebuild Managed System**.

After you select Rebuild Managed System, the current system information appears.

3.2.3 Managing a frame and resources connected to the HMC

A frame is a collection of a managed system and resources. Each frame is shown in the Contents area as the root of a resource tree. The managed system is listed underneath the frame.

If the managed system does not appear under the frame in the Contents area, refresh the frame as follows:

1. In the Contents area, select the frame.
2. From the menu, select **Selected**.
3. Select **Rebuild Managed Frame**.

Any user role can refresh a frame. Refreshing the frame is the only option available at the frame level.

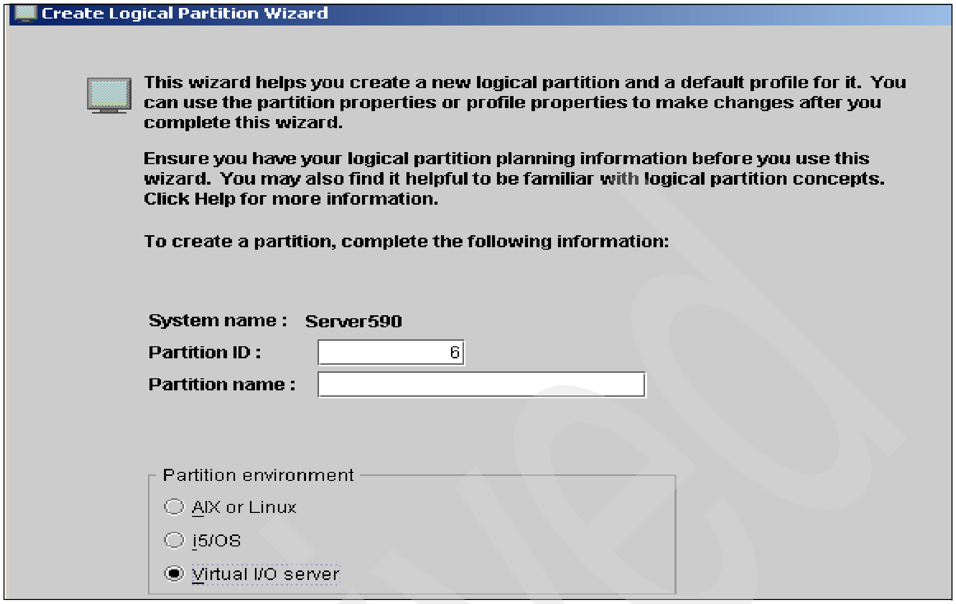
3.3 Logical partitioning of a system

Partitioning your system is similar to partitioning a hard drive. When you partition a hard drive, you divide a single physical hard drive so that the operating system recognizes it as a number of separate logical hard drives. You have the option of dividing the system's resources (memory, processor, or adapters) by using the HMC to partition your system. On each of these partitions, you can install an operating system and use each partition as you would a separate physical system. With the introduction of Virtual I/O Server, it is possible to further partition I/O adapters as well as processors.

3.3.1 Types of partitions

The HMC enables you to use the following types of partitions: logical partitions, Micro-Partitioning, and the Full System Partition. We introduce Micro-Partitioning in Chapter 6, "Advanced POWER Virtualization" on page 191.

The most important features to know about virtualization and Micro-Partitioning include the ability to partition a processor and share I/O adapters on a single system. Each system will have its own operating system and act independently. A Virtual I/O Server logical partition needs to be created before other partitions can use virtual I/O because it controls the resources shared between all other LPARs. Figure 3-3 on page 78 shows the creation of a Virtual I/O Server.


Create Logical Partition Wizard

This wizard helps you create a new logical partition and a default profile for it. You can use the partition properties or profile properties to make changes after you complete this wizard.

Ensure you have your logical partition planning information before you use this wizard. You may also find it helpful to be familiar with logical partition concepts. Click Help for more information.

To create a partition, complete the following information:

System name : Server590

Partition ID : 6

Partition name :

Partition environment

○ AIX or Linux

○ i5/OS

● Virtual I/O server

Figure 3-3 The selection for creating a Virtual I/O Server

Logical partition

Logical partitions are user-defined system resource divisions. Administrators determine the number of processors, memory, and I/O that a logical partition can have when active.

Full System Partition

A Full System Partition assigns the managed system's resources to one large partition. The Full System Partition is similar to the traditional, non-partition method of operating a system. You can use AIX 5L commands such as **lsdev** and **lscfg** to view resources. Because all resources are assigned to this partition, no other partitions can be started when the Full System Partition is running. Likewise, the Full System Partition cannot be started while other partitions are running.

The HMC enables you to easily switch from the Full System Partition to logical partitions. The setup of the operating system in a partition might require some careful planning to ensure that no conflicts exist between the two environments.

3.3.2 Managing a partitioned system

The HMC can be used to manage your partitioned system. Different managed-object types exist within the user interface. You can perform management functions by selecting the appropriate object type and then selecting an appropriate task. The main types of objects are managed systems, partitions, and profiles.

Managed systems

Managed systems are systems that are physically attached to and managed by the HMC. The HMC can perform tasks that affect the entire managed system, such as powering the system on and off. You can also create partitions and profiles within each managed system. These partitions and profiles define the way you configure and operate your partitioned system.

Partitions

Within your managed system, you can assign resources to create partitions. Each partition runs a specific instance of an operating system. The HMC can perform tasks on individual partitions. These tasks are similar to those you can perform on traditional, non-partitioned servers. For example, you can use the HMC to start the operating system and access the operating system console.

Because the HMC provides a virtual terminal for each partition, a terminal window can be opened for each partition. This virtual terminal can be used for software installation, system diagnostics, and system outputs. Only one instance of the terminal can be opened at a time. The managed system firmware and device drivers provide the redirection of the data to the virtual terminal.

Profiles

A profile defines a configuration setup for a managed system or partition. You can then use the profiles you created to start a managed system or partition in a particular configuration.

3.3.3 Partition profiles

A partition does not actually own any resources until it is activated. Resource specifications are stored within partition profiles. The same partition can operate using different resources at different times, depending on the profile you activated. When you activate a partition, you enable the system to create a partition using the set of resources in a profile created for that partition. For example, a logical partition profile might indicate to the managed system that its partition requires three processors, 2 GB of memory, and I/O slots 6, 11, and 12 when activated.

You can have more than one profile for a partition. However, you can only activate a partition with one profile at a time. Partition profiles are not affected by changes you make using the dynamic logical partitioning feature. If you want permanent changes, you must then reconfigure partition profiles manually. For example, if your partition profile specifies that you require two processors and you use dynamic logical partitioning to add a processor to that partition, you must change the partition profile if you want the additional processor to be added to the partition the next time you use the profile.

3.3.4 System profiles

Using the HMC, you can create and activate often-used collections of predefined partition profiles. A collection of predefined partition profiles is called a system profile. The system profile is an ordered list of partitions and the profile that is to be activated for each partition. The first profile in the list is activated first, followed by the second profile in the list, followed by the third, and so on.

The system profile helps you change the managed systems from one complete set of partitions configurations to another. For example, a company might want to switch from using 12 partitions to using only four every day. To do this, the system administrator deactivates the 12 partitions and activates a different system profile, one specifying four partitions. You cannot select a system profile from the power-on screen when there are no predefined system profiles. The option appears unavailable.

3.3.5 Overall requirements for partitioning

Before you begin to create partitions, complete the following activities:

- Record the required subnet mask, any gateway information, and address of your DNS server.
- Check that you have a suitable LAN (hub or switch and cables) to connect to each HMC and each network adapter used by partitions.
- Record the TCP/IP names and addresses to be resolved by a DNS server, or to be entered into the /etc/hosts file in each partition, and on the HMC.
- Determine the following information:
  - Your current resources for each partition
  - The operating system host name for each partition
  - The partition you want to use for service actions
  - The operating system to be loaded on the partition

3.4 LPAR resource administration

Allocating resources, changing partition profiles, and modifying system profiles are key server management tasks. These tasks require special levels of authority and are very significant in LPAR administration.

3.4.1 Assignable resources for logical partitioning

A partition profile stores the information of the three types of resources: CPU, memory, and I/O slots. Partition profiles store the information of the specific PCI slots assigned in the I/O drawers where the I/O devices are possibly plugged in.

> **Note:** Partition profiles do not store the information about specific I/O devices and PCI adapters.

Processors

Each installed and configured processor in the partioning-capable pSeries server can be assigned to a partition. You do not have to specify the precise location of the assigned processors in the partition profile, because the system selects the resources automatically. At least one processor must be assigned to each partition. Starting with the POWER5 technology, sharing processors between multiple active partitions is possible.

Memory

In a partitioned environment, some of the physical memory areas are reserved by several system functions to enable partitioning in the partioning-capable pSeries server. You can assign unused physical memory to a partition. The minimum amount of physical memory for each partition is 128 MB. You can assign further physical memory to partitions in increments of 128 MB.

I/O slots

I/O devices are assignable to partitions on a PCI slot (physical PCI connector) basis. This means that it is not the PCI adapters in the PCI slots that are assigned as partition resources, but the PCI slots into which the PCI adapters are plugged. To install an operating system, you need a device adapter and a boot/install media. For application and network usage, you also need a network adapter. The adapters and boot disks may or may not be physical adapters. Refer to Chapter 6, “Advanced POWER Virtualization” on page 191 to learn more about virtual adapters.

Types of values for resource assignment

In a partition profile, you need to specify three types of values for each resource. For CPU and memory, specify *minimum*, *desired*, and *maximum* values.

If any of the three types of resources cannot satisfy the specified minimum and required values, the activation of a partition will fail. If the available resources satisfy all the minimum and required values, but do not satisfy desired values, the activated partition will get as many of the resources as are available.

The maximum value is used to limit the maximum CPU and memory resources when dynamic logical partitioning operations are performed on the partition. If you exceed the total amount of memory in the system during profile creation, an error occurs stating that you have exceeded maximum system memory.

> **Note:** The maximum value is not shown and is unavailable if the HMC software level is earlier than Release 3, Version 1.

If you are going to install operating systems that do not support dynamic logical partitioning, you should specify the same values for both the desired and maximum values.

Assigning a host name to a partition

Each partition, including the Full System Partition, must have a unique host name that can be resolved. Host names cannot be reused between the Full System Partition and the logical partitions. If you need to change the host name of the partition manually, you might need to update the Network Settings on the HMC. You need to update if a short partition name is used or if a DNS server is not used.

3.4.2 Operating states

The HMC Contents area displays an operating state for the managed system.

Operating states for managed systems

The operating states provided in Table 3-2 apply to the managed system.

Table 3-2 States for managed systems

State	Description
Operating	The managed system is powered on and is initializing.
Power off	Server is off.
Initializing	The managed system is powered on and is initializing.

State	Description
Pending authentication	System is waiting for password to be authenticated.
Failed authentication	Processor password not synchronized with HMC.
Error	The managed system's operating system or hardware is experiencing errors.
On demand recovery	System VPD card is replaced and waiting replacement code for Virtual engine or Capacity on Demand technologies.
No connection	The HMC cannot contact the managed system.
Standby	System is on LPAR mode and awaits LPARs to be restarted.
Incomplete	HMC failed to get information from the managed system.
Recovery	Save area in the service processor is not synchronized with the HMC database.

Operating states for partitions

The operating states provided in Table 3-3 apply to the logical partition.

Table 3-3 States for partitions

State	Description
Ready	The partition is not active but is ready to be activated.
Starting	The partition is activated and is booting.
Running	The partition has finished its booting routines. The operating system might be performing its booting routines or is in its normal running state.
Error	The managed system's operating system or hardware is experiencing errors.
Not available	This partition is not available for use. Reasons can include: ► The managed system is powered off. ► The Full System Partition is not available when the managed system is powered on with the Partition Standby power-on option. ► Logical partitions are not available when the managed system is powered on with the Full System Partition power-on option.
Shutting down	LPAR is powering off.
Open firmware	The partition was activated by a profile that specified an OPEN_FIRMWARE boot mode.

3.4.3 Creating logical partitions

To create a logical partition, perform the following steps:

1. In the Contents area, select the managed system.
2. From the Selected menu, select **Create**.
3. Select **Logical Partition**. The Create Logical Partition and Profile wizard opens. See Figure 3-3 on page 78 for logical partition creation.
4. In the first window of the Create Logical Partition and Profile wizard, provide a name for the partition profile that you are creating. Use a unique name for each partition that you create. Names can be up to 31 characters long.

   Click **Next**.
5. Type the name of the profile you are creating for this partition.

   Click **Next** unless you want to assign the partition to a workload management group.
6. Select the desired, minimum, and maximum number of processors you want for this partition profile.

   Click **Next**.
7. Select the desired and minimum number of memory.

   Click **Next**.
8. The left side of the new window displays the I/O drawers available and configured for use. To expand the I/O tree to show the individual slots in each drawer, click the icon next to each drawer. Because the HMC groups some slots, if you attempt to assign a member of one of these grouped slots to a profile, the entire group is automatically assigned. Groups are indicated by a special icon named Group_XXX.

   Select the slot for details about the adapter installed in that slot. When you select a slot, the field underneath the I/O drawer tree lists the slot's class code and physical location code.

   > **Note:** The slots in the I/O Drawers field are not listed in sequential order.
9. Select the slot you want to assign to this partition profile and click either **required** or **desired**. If you want to add another slot, repeat this process. Slots are added individually to the profile; you can add slots one at a time, unless they are grouped. Minimally, assign a boot device to the **required** list box. This might be real or virtual depending on the choice of your server.

   There are two groups to which you can add adapters: a *required* group and a *desired* group. Desired adapters will be used if they are available at the time of activation. Required adapters are adapters that you require for this partition. If the adapters in this group are not available at the time of activation, the partition does not activate.

Click **Next**.

10. This window enables you to set service authority and boot mode policies for this partition profile.

    Select the boot mode that you want for this partition profile. Click **Next**.

11. This window supplies you with summary information about this partition.

    Review the information to ensure that you have the appropriate resources assigned to this partition.

12. If you want to change the configuration, click **Back**. Otherwise, click **Finish** to create the partition and profile.

13. The new partition, along with the default profile you just created appears under the Managed System tree in the Contents area.

    After you have created a partition, you must install an operating system and configure inventory data collection on the HMC and on the partition.

    > **Note:** For operating system installation steps, refer to 3.5, “Installation of the operating system into a partition” on page 92.

3.4.4 Activating partitions

To activate a partition, select the partition itself, and click **Activate** from the Selected menu. A window opens that lists activation profiles. The default partition profile is automatically highlighted, but you can activate the partition with any of the listed profiles.

If the required resources you specified in the partition profile that you are using to activate the partition exceed the amount of available resources, this partition does not activate. All resources currently not being used by active partitions are considered available resources. It is important that you keep track of your system’s resources at all times.

> **Note:** The `lshsc` command (using SSH) can be used to determine available resources as long as all the LPARs are in the running state.

For service, you must also configure Inventory Scout Services for each partition you activate.To activate partitions, you must be a member of one of the following roles: Operator, Advanced Operator, System Administrator, or Service Representative.

Activating a specific partition profile

To activate a partition profile, perform the following steps:

1. In the Contents area, select a partition profile.
2. From the menu, select **Selected**.
3. Select **Activate**.
4. The profile name is highlighted. Click **OK** to activate the partition profile.

The virtual operator panel next to the partition cycles through hardware boot sequence error and information codes, and then displays operating system error and information codes. For a complete description of these codes, refer to the hardware service documentation provided with your managed system and the documentation provided with your operating system.

Activating the default partition profile

To activate a partition without selecting a specific partition profile, perform the following steps:

1. In the Contents area, select the partition.
2. From the menu, select **Selected**.
3. Select **Activate**.
4. The default profile name is highlighted. Click **OK**. If you want to activate a different profile, select another profile in the list and then click **OK**.

Reactivating a partition with a partition profile

Reactivating a partition with a different profile requires shutting down the operating system that is running in that partition and activating another profile.

To reactivate a partition with a partition profile, you must be a member of one of the following roles: Operator, Advanced Operator, System Administrator, or Service Representative.

To reactivate a partition with a different profile, perform the following steps:

1. In the Contents area, select the partition for which you want to change profiles.
2. Open a terminal window for that partition to look at the operating system.
3. Run an appropriate `shutdown` command. The system shuts down the operating system, and the partition's state changes from Running to Ready in the Contents area. The `shutdown -Fr` command will not change the hardware configuration of a profile.

4. In the Contents area, select the new partition profile you want to activate for that partition.
5. From the menu, select **Selected**.
6. Select **Activate**.

3.4.5 Deleting partitions

To delete a partition, the managed system must be powered on using the Partition Standby power-on option. If you delete a partition, all of the profiles associated with that partition are also deleted. The partition is also automatically deleted from all system profiles.

You can delete partitions if you are a member of the System Administrator role. You cannot delete an activated partition. The option will be unavailable.

To delete a partition, perform the following steps:

1. Select the partition from the Contents area.
2. From the menu, select **Selected**.
3. Select **Delete**.

3.4.6 Partition startup and shutdown

This section describes the processes for starting and stopping partitions containing the AIX 5L Version 5.3 operating system. We cover the following topics:

- Normal partition startup
- Partition shutdown
- Partition reboot
- Partition recovery
  - Soft reset
  - Hard reset

Normal partition startup

To start AIX 5L in a partition, make sure that the partition profile boot mode is set to NORMAL and the partition is in the Ready state. Ensure that the default profile has sufficient resources to start the operating system.

To activate the partition, select the partition, right-click, and select **Activate**. When you select Activate, the HMC uses the boot mode selection in the partition profile.

The partition performs a virtual power-on, and the operating system starts booting. The partition then enters the Starting state, and the LED codes appear on the virtual operator panel within the HMC.

Partition shutdown

When a partition is up, it is in the Running state. To use AIX 5L to shut down an LPAR, perform one of the following actions:

- At the AIX 5L command prompt, type:

  ```
  shutdown -F
  ```

- For a forced shutdown, right-click the partition and select **Shut Down Partition**.

The partition eventually changes to the Ready state. You have now shut down AIX 5L and its partition.

Partition reboot

When a partition is up, it is in the Running state. To use AIX 5L to reboot an LPAR, type the following command at the AIX 5L command prompt:

```
shutdown -Fr
```

Notice the use of the `-r` flag; this tells the system to reboot after shutting down. The partition eventually changes to the Starting state, and AIX 5L begins booting. The virtual operator panel displays LED codes during the boot.

Note that this will not activate resources that have been added to a profile. The partition must be shut down to the Ready state, and the changes must be activated using the HMC.

Partition recovery

When an operating system stops responding, the HMC enables the operating system on a partition to be reset when errors are encountered in the operating system. The system can undergo either a soft or hard reset, as described in the following sections.

Soft reset

Soft reset actions are determined by your operating system's policy settings. Depending on how you have configured these settings, the operating system might do the following actions:

- Perform a dump of system information
- Restart automatically

If there is a problem with the running operating system, the following procedure performs a soft reset. This is the equivalent of pressing the yellow reset button on an RS/6000 system.

To reset the operating system on a partition, select the partition, right-click, and select **Restart partition**. Choose between the following options: Dump, operating systems and Operating system immediate. Figure 3-4 on page 90 shows the Reset Partition window.

In the window that opens, select **soft reset** and click **Yes**.

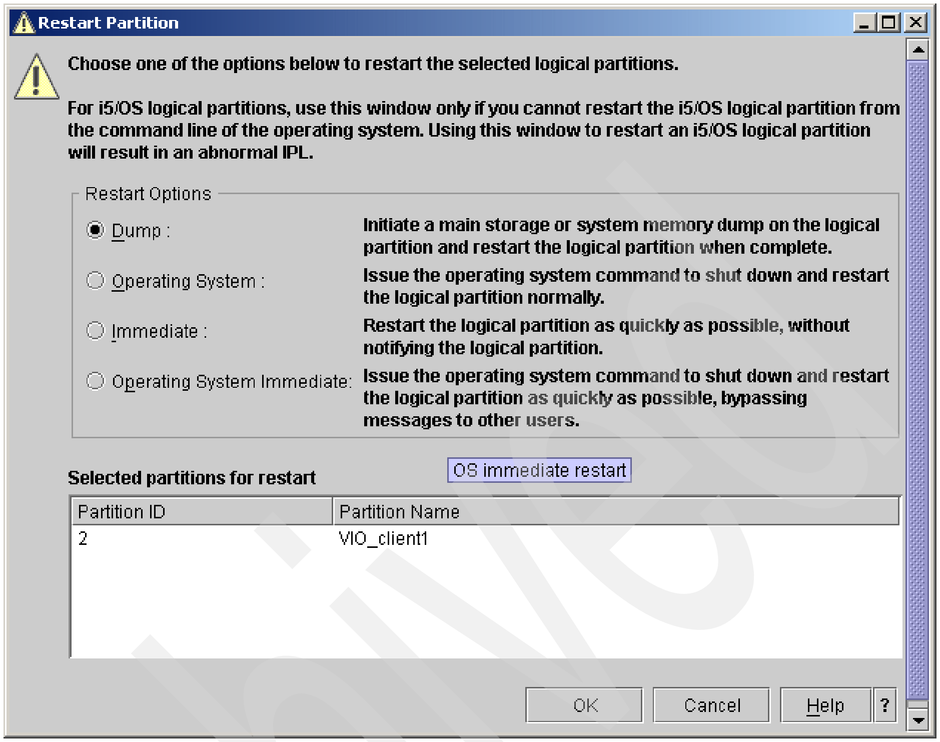
Hard reset

A hard reset virtually powers off the system. Issuing a hard reset forces termination and can corrupt information. Use this option only if the operating system is disrupted and cannot send or receive commands. To reset the operating system on a partition, select the partition, right-click, and select **Reset**. Select **Operating System**. Figure 3-4 on page 90 shows the Reset Partition window.

In the window that opens, select **hard reset** and click **Yes**.

The partition powers off and returns to a Ready state.

> **Important:** Pay attention to the message about other operating systems, such as Linux and i5/OS, when selecting reset. See Figure 3-4.



Restart Partition

Choose one of the options below to restart the selected logical partitions.

For i5/OS logical partitions, use this window only if you cannot restart the i5/OS logical partition from the command line of the operating system. Using this window to restart an i5/OS logical partition will result in an abnormal IPL.

Restart Options

● Dump : Initiate a main storage or system memory dump on the logical partition and restart the logical partition when complete.

○ Operating System : Issue the operating system command to shut down and restart the logical partition normally.

○ Immediate : Restart the logical partition as quickly as possible, without notifying the logical partition.

○ Operating System Immediate: Issue the operating system command to shut down and restart the logical partition as quickly as possible, bypassing messages to other users.

Selected partitions for restart

Partition ID	Partition Name
2	VIO_client1

Figure 3-4 Options to reset partitions

3.4.7 Setting up service authority

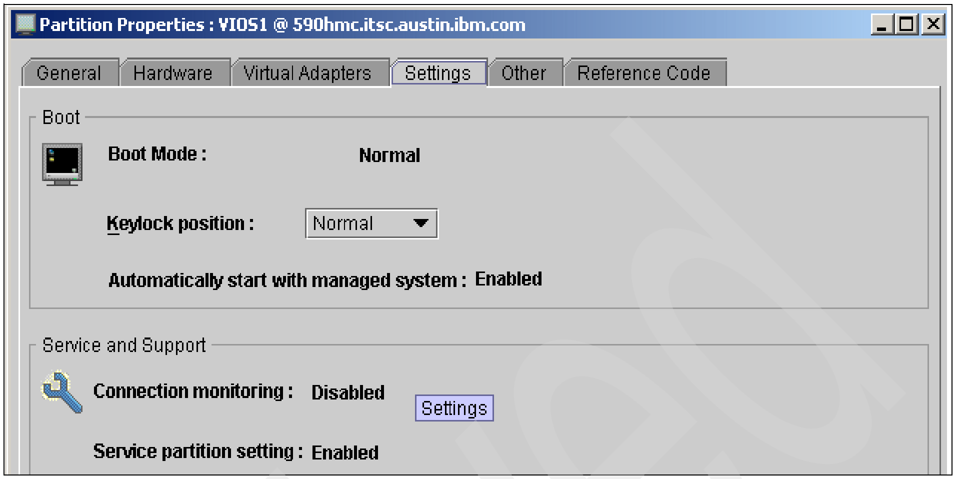
When partitions are created, we recommended that one partition is given service authority. The partition designated as having service authority can be used to perform system firmware upgrades and set other system policy parameters without having to power off the managed system.

The partition with the service authority must have a floppy drive or a CD-ROM attached to it. The Virtual I/O Server is a perfect candidate for a service authorized partition.

To see if a partition has service authority enabled, perform the following steps:

1. Select the partition.
2. Right-click and select **Properties**.
3. Go to the Settings tab.

Figure 3-5 shows service authorized partition enabled.



Boot Mode : Normal

Keylock position : Normal

Automatically start with managed system : Enabled

Connection monitoring : Disabled

Service partition setting : Enabled

Figure 3-5 Service authorization partition enabled

3.4.8 Changing default partition profiles

When you create a partition, the HMC requires that you create at least one profile called the default profile. In the Contents area, the default profile is represented by an icon. The default profile cannot be deleted.

When activating a partition, the HMC highlights the default profile as the one to use during activation unless you specify that it activate a different partition profile.

To change default profiles, you must be a member of one of the following roles: System Administrator or Advanced Operator.

To change the default partition profile, perform the following steps:

1. In the Contents area, select the default partition profile that you want to change.
2. From the menu, select **Selected**.
3. Select **Change Default Profile**.
4. Select the profile that you want to make the default profile from the list.

3.4.9 Viewing system profile properties

Any user can view system profile properties.

To view the properties of the system profile, perform the following steps:

1. In the Contents area, select the system profile.
2. From the menu, select **Selected**.
3. Select **Properties**.

3.4.10 Modifying system profile properties

To modify system profiles, you must be a member of one of the following roles: System Administrator or Advanced Operator.

To modify system profiles, perform the following steps:

1. In the Contents area, select the system profile that you want to modify.
2. From the menu, select **Selected**.
3. Select **Properties** from the menu.
4. Change the system profile information as appropriate.

3.5 Installation of the operating system into a partition

This section provides an overview of the AIX 5L installation process and various installation scenarios you can use.

3.5.1 Operating system capability and configuration

The IBM @server p5 590 and p5 595 servers are capable of running the following operating systems:

- AIX 5L

  AIX 5L for POWER Version 5.2 with the 5200-04 Recommended Maintenance Package (APAR IY56722), or later, plus APAR IY60347 and AIX 5L for POWER Version 5.3 with APAR IY60349, or later

  > **Note:** Advanced POWER Virtualization is not supported on AIX 5L for POWER Version 5.2. It requires AIX 5L Version 5.3.

- Linux
  - Novell SUSE LINUX Enterprise Server 9 for POWER, or later
  - Red Hat Enterprise Linux AS for POWER Version 3, or later
- i5/OS

  i5/OS V5R3, or later

The p5-570 server is capable of running the following operating systems:

- AIX 5L

  AIX 5L Version 5.3 or AIX 5L Version 5.2 Maintenance Package 5200-04 (IY56722), or later
- Linux
  - SUSE LINUX Enterprise Server 9 for POWER systems, or later
  - Red Hat Enterprise Linux AS for POWER Version 3, or later

> **Note:** The following notes apply:
>
> - For Linux:
>   - Not all server features available on the AIX 5L operating system are available on the Linux operating systems.
>   - Dynamic LPAR is not supported by Red Hat Enterprise Linux AS for POWER Version 3.
>   - IBM only supports the Linux systems of clients with a SupportLine contract covering Linux. Otherwise, contact the Linux distributor for support.
> - For i5/OS:
>   - It is supported on 1.65 GHz POWER5 models only.
>   - Only one or two processors on the p5-590 and p5-595 systems can be dedicated to i5/OS, while the other servers do not support i5/OS at the time of writing.
>   - Only an AIX 5L or Linux logical partition can be designated as the service partition. An i5/OS partition cannot be designated as the service partition.

3.5.2 Recommended configuration for a basic AIX 5L installation

When installing your AIX 5L system, it is important to understand the basic usage of the system. This enables you to adequately design a system capable of handling its workload.

Consider the following list of recommendations for a basic installation:

- Hard disks

  Use at least two physical volumes (PV) to be used by rootvg. Each of the volumes needs to be able to handle the operating system on its own. These physical volumes need to be mirrored. The operating system is not more than 3 GB in size, but 5 GB advised for future growth, maintenance upgrades, and fixes.

- File systems

  Some applications use root file systems as their root for binaries, and this can be hard coded. Try to isolate the binaries from the root file system by using a separate file system with a mount point pointing to a root file system. For example, if an application uses a directory such as /usr/tivoli, create a file system with a mount point of /usr/tivoli. In doing that, /usr becomes a file system and /usr/tivoli another file system.

- Paging space

  This is controlled mainly by the applications running on the system. For example, some applications require that your paging space be three to four times the size of your real memory. AIX 5L provides the following default sizes based on your initial configuration.

  Paging space can use no less than 16 MB, except for hd6, which can use no less than 64 MB in AIX Version 4.3 and later. Paging space can use no more than 20% of total disk space. If real memory is less than 256 MB, paging space is twice the real memory. If real memory is greater than or equal to 256 MB, paging space is 512 MB.

- System dump device

  After system installation, use the **`sysdumpdev -e`** command to estimate the size of the system dump device. AIX 5L Version 5.3 isolates the dump device from the default hd6 to lg_dumplv. Make sure that **`dumpcheck`** is activated in cron. It periodically estimates the size of the system dump and compares it with the size of the dump device. It will notify the administrator when the dump device becomes too small.

- Memory

  This depends on the application. The more the memory you have, the more chances of improving performance. This is only a subset of performance factors. The whole set is beyond the scope of this publication. *AIX 5L Practical Performance Tools and Tuning Guide*, SG24-6478, provides more information about system performance.

- Network

  If possible, we recommend two or more physical adapters for both redundancy and load balancing. You might need a separate adapter to run backups.

3.5.3 Disks, boot devices, and media devices

The servers and partitions must have access to a device capable of reading CD media or to a Network Installation Management (NIM) server. A minimum of two internal SCSI hard disks are required per server, but not for the LPARs. We recommend that you use these disks as mirrored boot devices. Mount these disks in the first I/O drawer whenever possible.

Boot support is available from local SCSI, virtual SCSI, SSA, and Fibre Channel adapters, or from networks using Ethernet, virtual Ethernet, or token-ring adapters.

Virtual I/O:

- Virtual Ethernet clients can only connect partitions within a single system.
- The target partition must run AIX 5L Version 5.3 or Linux with the appropriate kernel or a kernel that supports virtualization. AIX 5L V5.2 does not support virtualization.
- Virtual disk and virtual Ethernet depend on the availability of the Virtual I/O Server.

For more information about virtual I/O, refer to 6.2, “Micro-Partitioning” on page 193.

3.5.4 Overview of the AIX 5L installation process

Due to the physical configuration of a managed system, we recommend the use of the Network Installation Management (NIM) environment to install AIX 5L. Table 3-4 on page 96 compares how different forms of media proceed through the AIX 5L installation process.

Table 3-4 AIX 5L installation process comparison

Steps	CD-ROM produce media	NIM	mksysb on CD-R or DVD-RAM	mksysb on tape
Booting	Boot image is stored and retrieved from CD.	Boot image is stored on NIM server and retrieved from network by firmware.	Boot image is stored and retrieved from CD-R or DVD.	Boot image is stored and retrieved from the first image on tape.
Making BOS installation choices	Manually step through the BOS menu selections for disks, kernel, language, and so on.	Perform nonprompted installation using a bosinst.data file to answer the BOS menu questions.	Manually proceed through the BOS menu selections for disks.	Manually proceed through BOS menu selections for disks and other choices.
Executing commands during installation	CD-file system is mounted, and commands are executed.	Shared Product Object Tree (SPOT) file system is NFS mounted, and commands are run from the SPOT.	CD-file system is mounted, and commands are executed.	Command files are retrieved from second image on tape to RAM file system in memory.
Installing product images	Installation images stored on CD in a file system.	Installation images are stored in LPP_Source, which is NFS mounted during installation.	Backup image is stored on CD-R or DVD-RAM in a file system.	Backup image is stored and retrieved from fourth image on tape.
Rebooting system and logging in to the system	Use configuration assistant (or installation assistant) to accept license agreements, set paging space, and so on.	No configuration assistant (or installation assistant). Boot to login prompt.	No configuration assistant (or installation assistant). Boot to login prompt.	No configuration assistant (or installation assistant). Boot to login prompt.

3.5.5 Introduction to Network Installation Management

Network Installation Management (NIM) gives you the ability to install and maintain the AIX 5L operating system (BOS) and any additional software and fixes that might be applied over time. It can be used to customize the configuration of machines both during and after installation. Using NIM eliminates the need for access to physical media, such as tapes and CD-ROMs, because the media is a NIM resource on a NIM server. System backups can be created with NIM and stored on any server in the NIM environment, including the NIM master. Use NIM to restore a system backup to the same partition or to another partition. Before you begin configuring the NIM environment, you must have completed the following actions:

- NFS and TCP/IP installed
- TCP/IP configured correctly
- Name resolution configured

The nimish command

In previous versions of AIX, NIM used the **rsh** and **rcmd** commands to perform remote execution of commands on clients. These r-commands were a potential security exposure.

AIX 5L Version 5.3 is enhanced with the **nimsh** command environment that is part of the bos.sysmgt.nim.client fileset.

The AIX 5L **which_fileset** command is a useful tool that searches the /usr/lpp/bos/AIX_file_list file for a specified file name or command.

On NIM operations, the client and server send authentication information that will be validated by **nimish**. The **nimish** command requires the following authentication information:

- Host name of NIM client
- CPUID of NIM client
- CPUID of NIM master
- Return port for secondary (stderr) connection
- Query flag

The **rsh** and **rcmd** commands are still supported for compatibility reasons. The **nimish** command allows the following two remote execution environments:

- NIM service handler for client communication: Basic **nimsh**
- NIM cryptographic authentication: OpenSSL

Existing clients not using the current **nimish** command authentication can be updated using the **smitty nim_config_services** fast path. Although it is not recommended, the use of the **niminit** command will archive the same goal. Example 3-2 on page 98 shows the execution of the **niminit** command.

Example 3-2 Using niminit for using nimish

```
# . /etc/niminfo
# mv /etc/niminfo /etc/niminfo.bak
# niminit -aname=server2 -amaster=server4 -aconnect=nimsh
nimsh:2:wait:/usr/bin/startsrc -g nimclient >/dev/console 2>&1
0513-044 The nimsh Subsystem was requested to stop.
0513-059 The nimsh Subsystem has been started. Subsystem PID is 442484.
# nimclient -C
0513-059 The nimsh Subsystem has been started. Subsystem PID is 417820.
```

OpenSSL

AIX 5L Version 5.3 introduces a cryptographic extension for NIM on the OpenSSL environment, which is delivered on the Linux Toolbox for AIX media.

High Available NIM (HA_NIM)

The most significant single point of failure in a NIM environment is the NIM master. AIX 5L Version 5.3 introduces a way to define a backup NIM master to take over to the backup master and then fall back to the primary master. The takeover is done by an administrator command from the backup server. Figure 3-6 shows a typical HA_NIM environment. The NIM master has to run a backup operation regularly to synchronize the configurations from the primary to the backup server.

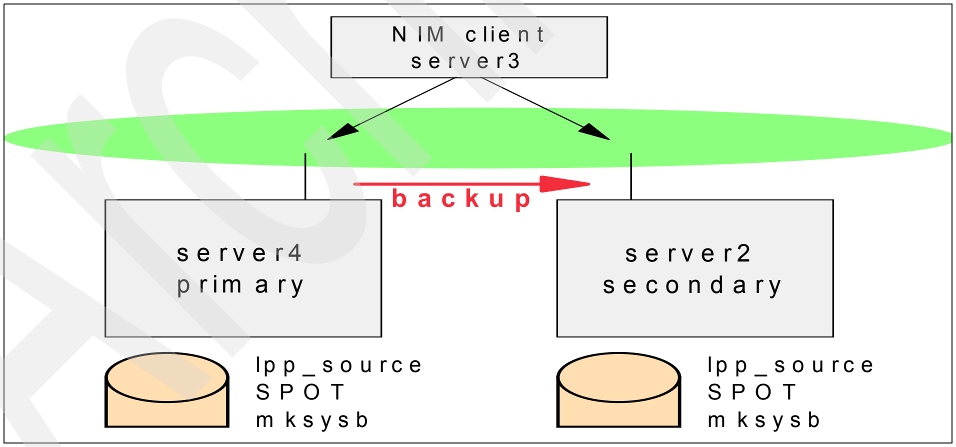


Figure 3-6 HA_NIM environment

To set up an HA_NIM environment, perform the following steps:

1. On the backup server, install the NIM master and Shared Product Object Tree (SPOT) filesets using regular AIX 5L procedures.

Define the primary master by running the **smit nim_mkaltmstr** command on the master or by using the following command:

```
nim -o define -t alternate_master -aplatform=chrp -aif1='network1 server2
0' \ -anetboot_kernel=mp server2.
```

2. Initialize the backup master:

```
# niminit -ais_alternate=yes -aname=server2 -amaster=server4 -apif_name=en0
\  -aplatform=chrp
```

3. Synchronize the NIM masters, by running:

```
nim_master_recover
```

Or:

```
nim -o sync -aforce=yes server2
```

Creating a NIM master

To create a NIM master, perform the following steps:

1. Make sure that the correct version of the AIX 5L operating system has been installed on the master (AIX 5L Version 5.2 with the current maintenance level or AIX 5L Version 5.3 with the correct maintenance level for clients using virtualization).
2. Ensure that the NIM master has enough disk space for NIM resources and a CD-ROM assigned to it.
3. Verify that your network environment is already defined and working correctly before configuring the NIM environment.
4. As the root user, you will set up the NIM environment using the **nim_master_setup** script. The **nim_master_setup** script automatically installs the bos.sysmgt.nim.master fileset, configures the NIM master, and creates the required resources for installation, including a mksysb system backup.

The **nim_master_setup** script uses the rootvg volume group and creates an /export/nim file system by default.

Using the NIM master to install partitions on the same system

To prepare a client LPAR for installation, perform the following steps:

1. Use the **nim_clients_setup** script to define your NIM clients.
2. Allocate the installation resources.
3. Initiate a NIM BOS installation on the clients.
4. Using the HMC, activate the client partitions and configure them to boot off the network.

Prerequisites

Before beginning with this procedure, the following tasks must be completed:

- A NIM master must be setup.
- Network communication to the master and the server must be configured.
- Use the HMC to create logical partitions and partition profiles for each NIM client. Be sure that each LPAR has a network adapter assigned. Set the boot mode for each partition to SMS mode. After you have successfully created the partitions and partition profiles, leave the partitions in the Ready state. Do not activate the partitions yet.

> **Note:** It is possible to use a virtual network with the NIM master and clients on the same system. If the LPARs use the virtual I/O, it is important to make sure that the Virtual I/O Server is operational, because there is no “real” network adapter between the NIM client and master.

Use the NIM master to perform the following steps:

1. Activate the Master_LPAR (using the HMC interface).
2. Configure the NIM master and initiate the installation of the partitions. Confirm or enter your host name, IP address, name server, domain name, default gateway, and ring speed or cable type when configuring your network.
3. Activate and install the partitions (using the HMC interface).

   To activate the partitions, perform the following steps:

   a. Select the partition (or partition profile) you want to activate.

   b. Right-click the partition (or partition profile) to open the menu.

   c. Select **Activate**. The Activate Partition menu opens with a selection of partition profiles. Select a partition profile that is set to boot to the SMS menu.

   d. Select **Open terminal** at the bottom of the menu to open a virtual terminal (vterm) window.

   e. Select **OK**. A vterm window opens for each partition. After several seconds, the System Management Services (SMS) menu opens in the vterm window.

After the installation completes and the system reboots, the vterm window displays a login prompt.

Using NIM on a separate system to install each partition

In this procedure, a separate system running AIX 5L Version 5.3 is used as a NIM master and server to install logical partitions. It is assumed that the steps to create the NIM master are completed. As with any NIM environment, you must make sure that your network environment is already defined and working correctly. In this case, the availability of a physical network is *required*.

You then define your clients using SMIT or using the NIM clients.def file. Next, use the **nim_clients_setup** script to allocate the installation resources, and initiate a NIM BOS installation on the clients. Then using the HMC, you activate the partitions and configure them to boot off the network.

Prerequisites

Before you begin this procedure, you should have used the HMC to create partitions and partition profiles for each partition that you want to install. Be sure that each partition has a network adapter assigned. Set the boot mode for each partition to be SMS mode. After you have successfully created the partitions and partition profiles, leave the partitions in the Ready state. Do not activate the partitions yet.

Perform the following steps:

1. Configure the NIM master and initiate the installation of the partitions (perform these steps in the AIX 5L environment).
2. Activate and install the partitions (using the HMC interface).
3. Log in to your partition (perform this step in the AIX 5L environment).

When the installation has completed and the system has rebooted, the vterm window displays a login prompt.

> **Note:** You can also select the **nimsh** command for the BOS install operation. At the time of writing, there was no SMIT menu to do this. Use the **nim** command as follows:
>
> ```
> nim -o bos_inst -a connect=nimsh -a source=rte -a spot=spot530 \ -a
> lpp_source=aix530 server3
> ```

3.5.6 Installing a partition using alternate disk installation

An existing disk image can be cloned to another disk or disks without using NIM. However, you can choose to use NIM at a later time. With the appropriate filesets installed, SMIT panels can be used to help facilitate the **alt_disk_install** command.

When using **alt_disk_install** command to clone a system image to another disk, the **-O** option must be used to remove references in the object data manager (ODM) and device (/dev) entries to the existing system. The **-O** flag tells the **alt_disk_install** command to call the **devreset** command, which resets the device database. The cloned disk can now be booted as though it were a new system.

The following steps illustrate an example of this scenario:

1. Boot the managed system as a Full System Partition so that you have access to all the disks in the managed system.
2. Configure the system and install the necessary applications.
3. Run the **alt_disk_install** command to begin cloning the rootvg on hdisk0 to hdisk1, as follows:

   ```
   # /usr/sbin/alt_disk_install -O -B -C hdisk1
   ```

   The cloned disk (hdisk1) is named altinst_rootvg by default.
4. Rename the cloned disk (hdisk1) to alt1 so that you can repeat the operation with another disk:

   ```
   # /usr/sbin/alt_disk_install -v alt1 hdisk1
   ```
5. Run the **alt_disk_install** command again to clone to another disk and rename the cloned disk, as follows:

   ```
   # /usr/sbin/alt_disk_install -O -B -C hdisk2
   # /usr/sbin/alt_disk_install -v alt2 hdisk2
   ```
6. Repeat steps 3 through 5 for all of the disks that you want to clone.
7. Use the HMC to partition the managed system with the newly cloned disks. Each partition you create will now have a rootvg with a boot image.
8. Boot the partition into SMS mode. Use the SMS MultiBoot menu to configure the first boot device to be the newly installed disk. Exit the SMS menus and boot the system.

3.5.7 Using a CD-ROM to manually install a partition

In this procedure, you use either the system's built-in CD device or a virtual CD device assigned to the LPAR to perform a new and complete Base Operating System installation on a partition.

Prerequisites

Before using this procedure, create a partition and partition profile for the client using the HMC. Assign the SCSI (or vSCSI in the Virtual I/O Server client) bus controller attached to the CD-ROM device and enough disk space for the AIX 5L operating system to the partition. Set the boot mode for this partition to be SMS mode. After you have successfully created the partition and partition profile, leave the partition in the Ready state.

Manually installing a partition using a CD-ROM

Perform the following steps:

1. Activate and install the partition (using the HMC interface).

   To set advanced options, press 3 and then press Enter. The advanced options available in AIX 5L Version 5.3 enable you to change the desktop application and enable or disable the Trusted Computing Base option.

2. Manage your partition (perform this step in the AIX 5L environment).

When the installation completes and the system reboots, the vterm window displays a login prompt.

> **Note:** The following changes apply to the BOS installation:
>
> - The AIX 5L operating system previously contained both a uniprocessor 32-bit kernel and a multiprocessor 32-bit kernel. Effective with AIX 5L Version 5.3, the operating system supports only the multiprocessor kernel.
> - AIX 5L Version 5.2 is the last release of AIX that supports the uniprocessor 32-bit kernel.
> - Prior to AIX 5L Version 5.3, the default kernel was 32-bit, but with AIX 5L Version 5.3, the default kernel is 64-bit with a 32-bit option. Enhanced journaled file system (JFS2) is the default file system when installing the 64-bit kernel.
>
> The AIX 5L Version 5.3 32-bit multiprocessor kernel supports the following systems: RS/6000, eServer pSeries, p5, and OEM hardware based on the Common Hardware Reference Platform (CHRP) architecture, regardless of the number of processors.

3.5.8 Choosing the 64-bit or 32-bit kernel on installation

It can be very difficult for system administrators to decide which default kernel to use on installation. The decision is mostly affected by the applications that will run on the system. We recommend choosing a 64-bit kernel. This is supported by the following points:

- The 64-bit kernel introduced with AIX 5L Version 5.1 addresses bottlenecks that might have limited throughput on 32-way systems.

- The 64-bit kernel also improves scalability by allowing you to use larger sizes of physical memory, while the 32-bit kernel is limited to 96 GB of physical memory.
- The performance of 64-bit applications running on the 64-bit kernel on POWER5-based systems should be greater than, or equal to, the same application running on the same hardware with the 32-bit kernel.
- The 64-bit kernel allows 64-bit applications to be supported without requiring system call parameters to be remapped or reshaped.
- 32-bit applications can run on the 64-bit kernel with less than 5% performance degradation compared to the 32-bit call because of parameter reshaping.
- 64-bit applications from AIX 4.3 need to be recompiled in order to correctly run on AIX 5L.

3.5.9 Creating a system backup

Before any system migration, upgrade, or installation, we recommended that you perform a full system backup of the old system. A system backup, referred to as *mksysb*, is an image of the operating system with all fixes, upgrades, and configurations applied after the BOS installation.

The following list provides the recommended methods of system backup listed in order of preference:

- Using NIM
- Using the **`alt_disk_install`** command
- Creating a bootable CD/DVD media
- Creating a bootable tape

Note: Because of its flexibility, NIM is the recommended method to back up and reinstall your partitions. See 5.8, “Planning and performing backup and recovery” on page 180 for more information about backup and recovery.

3.6 Dynamic LPAR operation

A dynamic logical partition provides the ability to logically attach and detach a managed system’s resources to and from a partition’s operating system without rebooting.

3.6.1 Dynamically adding resources to a partition

This task enables you to add resources, such as processors, memory, and I/O slots, to a partition without rebooting the partition's operating system.

Processors

You can add up to the amount of free system processors (processors that are not assigned to a running partition) to a partition. You cannot exceed the maximum number specified in the partition's active profile.

Memory

You can only add up to the amount of free memory, or memory that is not assigned to a running partition. You also cannot exceed the maximum number specified in the partition's active profile.

Adapters

You can dynamically add any free adapters to the partition.

3.6.2 Dynamically moving resources between partitions

This task enables you to move resources, such as processors, memory, and I/O slots, from one partition to another without rebooting either partition's operating system.

Processors

You can move processors from one partition to another partition. You have to select appropriate values so that these two partitions satisfy the following conditions:

- The partition whose processors will be removed must be assigned more than or equal to the number of processors defined in the partition profile as the minimum value after the operation.
- The partition whose processors will be added must be assigned less than or equal to the number of processors defined in the partition profile as the maximum value after the operation.

Memory

You can move memory from one partition to another partition. You have to select appropriate values so that these two partitions satisfy the following conditions:

- The partition whose memory will be removed must be assigned more than or equal to the memory size defined in the partition profile as the minimum value after the operation.

- The partition whose memory will be added must be assigned less than or equal to the memory size defined in the partition profile as the maximum value after the operation.

Adapters

This task enables you to move a PCI I/O slot containing a PCI adapter from one partition to another without rebooting the partition's operating system. Before moving an I/O slot, you must log in to the operating system on the source partition and unconfigure the adapter and the I/O slot. The adapter that you are going to move must not be defined as Required in the current active partition profile on the source partition.

> **Note:** To ensure that Service Focal Point and Dynamic LPAR operations continue to function correctly, do not dynamically move the Ethernet adapter, which is used to communicate with the HMC.

3.6.3 Dynamically removing resources from a partition

This task enables you to remove resources, such as processors, memory, and I/O slots, from a partition without rebooting the partition's operating system.

Processors

You can remove processors from a partition without rebooting the partition's operating system. When you remove a processor, it is freed by the partition and available for use by other partitions.

> **Note:** The number of processors remaining after the removal operation cannot be less than the minimum value specified in this partition's active profile.

Memory

You can remove memory from a partition without rebooting the partition's operating system. When you remove memory, it is freed by the partition and available for use by other partitions.

> **Note:** The size of the memory remaining after the removal operation cannot be less than the minimum value specified in this partition's active profile.

Adapters

This task enables you to remove I/O slots, which can contain an adapter, from a partition without rebooting the partition's operating system. Before continuing with this task, you must use the partition's operating system to manually deconfigure each adapter that you want to remove. You cannot remove an adapter defined as Required in the current active partition profile.

4

Hardware management

The Hardware Management Console (HMC) uses its connection to the processor subsystem to perform various functions. These functions include creating and maintaining a multiple partition environment, detecting, reporting, and storing changes in hardware conditions, and acting as a Service Focal Point for service representatives to determine an appropriate service strategy.

This chapter includes the following sections:

- HMC
- Service applications
- Licensed Internal Code maintenance
- LPAR Validation Tool (LVT)

4.1 HMC

The Hardware Management Console is a dedicated workstation that controls managed systems, including IBM *e*server hardware, logical partitions, and Capacity on Demand. To provide flexibility and availability, there are different ways to implement HMCs, including the local HMC, remote HMC, redundant HMC, and the Web-based System Manager Remote Client. One HMC is capable of controlling multiple POWER5 processor-based systems.

> **Note:** At the time of writing, one HMC supports up to 48 POWER5 processor-based systems and up to 256 LPARs using the HMC Machine Code Version 4.4.

4.1.1 Local HMC

A local HMC is any physical HMC that is directly connected to the system it manages through a private service network. An HMC in a private service network is a DHCP[1] server from which the managed system obtains the address for its service processor and Bulk Power Controller.

4.1.2 Remote HMC

A remote HMC is a stand-alone HMC or an HMC installed in a 19-inch rack that is used to remotely access another HMC. A remote HMC can be present in an open network.

4.1.3 Redundant HMC

A redundant HMC manages a system that is already managed by another HMC. When two HMCs manage one system, those HMCs are peers and can be used simultaneously to manage the system.

4.1.4 Web-based System Manager Remote Client

The Web-based System Manager Remote Client is an application that is usually installed on a PC. You can then use this PC to access HMCs remotely. Web-based System Manager Remote Clients can be present in private and open networks. You can perform most management tasks using the Web-based System Manager Remote Client.

[1] DHCP stands for Dynamic Host Control Protocol.

The remote HMC and the Web-based System Manager Remote Client provide you with the flexibility to access your managed systems (including HMCs) from multiple locations using multiple HMCs.

4.1.5 HMC connectivity

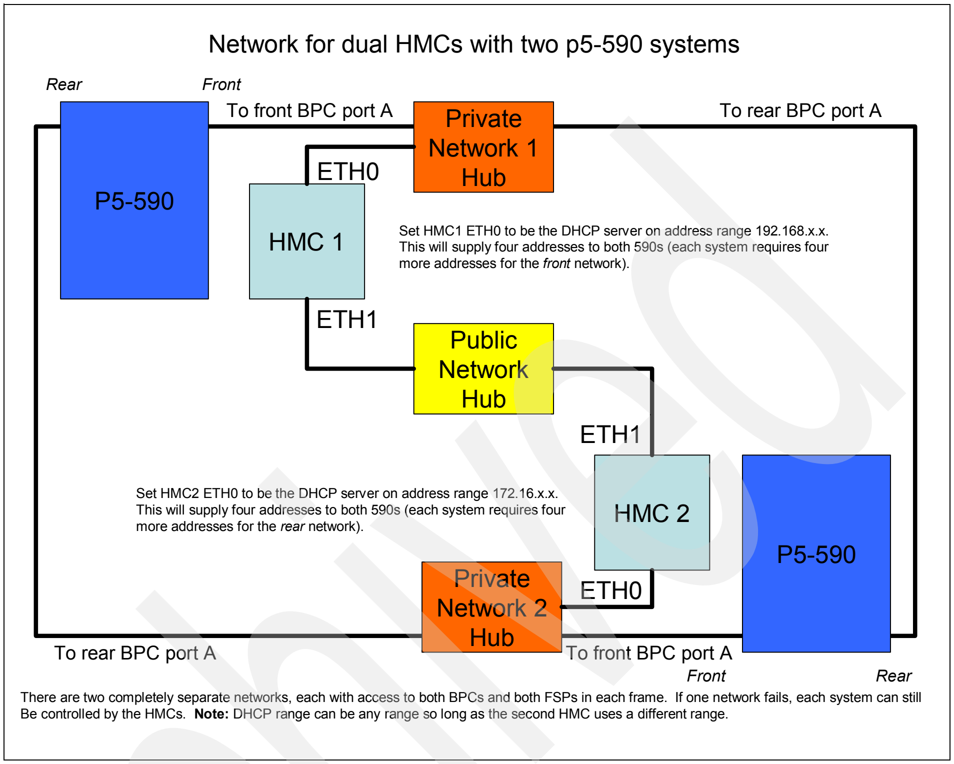
There are some significant differences regarding the HMC connection to the managed server with the p5-590 and p5-595 servers and other POWER5 processor-based systems:

- At least one HMC is mandatory, and two are recommended.
- The first (or only) HMC is connected using a private network to BPC-A (Bulk Power Controller). The HMC must be set up to provide DHCP addresses on that private (eth0) network.
- A secondary (redundant) HMC is connected using a separate private network to BPC-B. The second HMC must be set up as a DHCP server to use a different range of addresses for DHCP.
- Additional provision has to be made for an HMC connection to the BPC in a powered expansion frame.

> **Note:** DHCP must be used, because the BPCs depend on the HMC to provide them with addresses. There is no way to set a static address as with all other p5 systems on a BPC.

If there is a single managed server (with powered expansion frame), no additional LAN components are required. However, if there are multiple managed servers, an additional LAN switch or switches and cables will be needed for the HMC private networks. You must plan for these switches and cables.

Figure 4-1 shows two p5-590s controlled with dual HMCs.



Set HMC1 ETH0 to be the DHCP server on address range 192.168.x.x. This will supply four addresses to both 590s (each system requires four more addresses for the *front* network).

Set HMC2 ETH0 to be the DHCP server on address range 172.16.x.x. This will supply four addresses to both 590s (each system requires four more addresses for the *rear* network).

There are two completely separate networks, each with access to both BPCs and both FSPs in each frame. If one network fails, each system can still Be controlled by the HMCs. **Note:** DHCP range can be any range so long as the second HMC uses a different range.

Figure 4-1 Network for two p5-590s controlled with dual HMCs

HMC network interfaces

The HMC supports up to three separate physical Ethernet interfaces. In the desktop version of the HMC, this consists of one integrated Ethernet and up to two plug-in adapters. In the rack-mounted version, this consists of two integrated Ethernet adapters and up to one plug-in adapter. Use each of these interfaces in the following ways:

- One network interface can be used exclusively for HMC-to-managed system communications (and must be the eth0 connection on the HMC). This means that only the HMC, Bulk Power Controllers (BPCs), and service processors of the managed systems are on that network. Even though the network interfaces into the service processors are SSL-encrypted and password-protected, having a separate dedicated network can provide a higher level of security for these interfaces.

- Another network interface is typically used for the network connection between the HMC and the logical partitions on the managed systems for the HMC-to-logical partition communications.
- The third interface is an optional additional Ethernet connection that can be used for remote management of the HMC. This third interface can also be used to provide a separate HMC connection to different groups of logical partitions. For example, you can use any of the following options:
  - An administrative LAN that is separate from the LAN on which all the usual business transactions are running. Remote administrators can access HMCs and other managed units using this method.
  - Different network security domains for your partitions, perhaps behind a firewall with different HMC network connections into each of those two domains.

Note: With the rack-mounted HMC, if an additional (third) Ethernet port is installed in the HMC (using a PCI Ethernet card), that PCI-card becomes the eth0 port. Normally (without the additional card), eth0 is the first of the two integrated Ethernet ports.

4.1.6 HMC code

For updates of the machine code, HMC functions, BIOS updates, and hardware prerequisites, refer to the following Web page:

http://techsupport.services.ibm.com/server/hmc

POWER4 HMC models, such as the 7315-CR2 or 7315-C03, can be upgraded to support POWER5 processor-based systems.

BIOS updates

Before updating your HMC code, check that the correct BIOS level is installed.

Follow these steps to determine the current BIOS revision level on your HMC:

1. Shut down the HMC.
2. Power on the HMC, and then press the F1 key to display the Setup utility.
3. On the Main menu, check the Flash EEPROM Revision Level.

It is possible to start run the **rshterm** command on the HMC, or open a SSH connection to the HMC and enter following command:

```
hscroot@590hmc:~> lshmc -b
"bios=2AKT38RUS
```

> **Note:** It is not possible to connect POWER4 and POWER5 processor-based systems simultaneously to the same HMC.

To upgrade an existing POWER4 HMC:

- Contact your IBM Sales Representative for help.
- Call an IBM Service Center and order APAR MB00691.
- Order the CD online, selecting **Version 4.5 machine code updates** → **Order CD** → **Go** at the Hardware Management Console (HMC): Support for HMC for UNIX servers and Midrange servers Web page, available at:

  `http://techsupport.services.ibm.com/server/hmc`

> **Note:** You must have an IBM ID to use this freely available service. Registration information and online registration form is available at this Web page.

HMC fixes

HMC fixes are periodically released for the HMC. If you want to download HMC fixes from the service and support's system or Web site to your HMC or server, you must set up a connection to service and support either through a local or remote modem or through a VPN connection. You typically set up the service connection when you first set up your server. However, the service connection is not required for initial server setup.

If you do not have an Internet connection from the HMC, you must obtain the HMC fixes either on CD-ROM or on an FTP server. Refer to the following link to download or order HMC fixes:

`http://techsupport.services.ibm.com/server/hmc`

Backup the HMC

Before installing fixes and after changing LPARs or any configuration data you must backup the HMC.

When backing up critical HMC data using the HMC, you can backup all important data, such as:

- User-preference files
- User information
- HMC platform configuration files
- HMC log files

The Backup function saves the HMC data stored on the HMC hard disk to DVD, a remote system mounted to the HMC file system (such as NFS), or a remote site through FTP. Back up the HMC after you have made changes to the HMC or to the information associated with logical partitions.

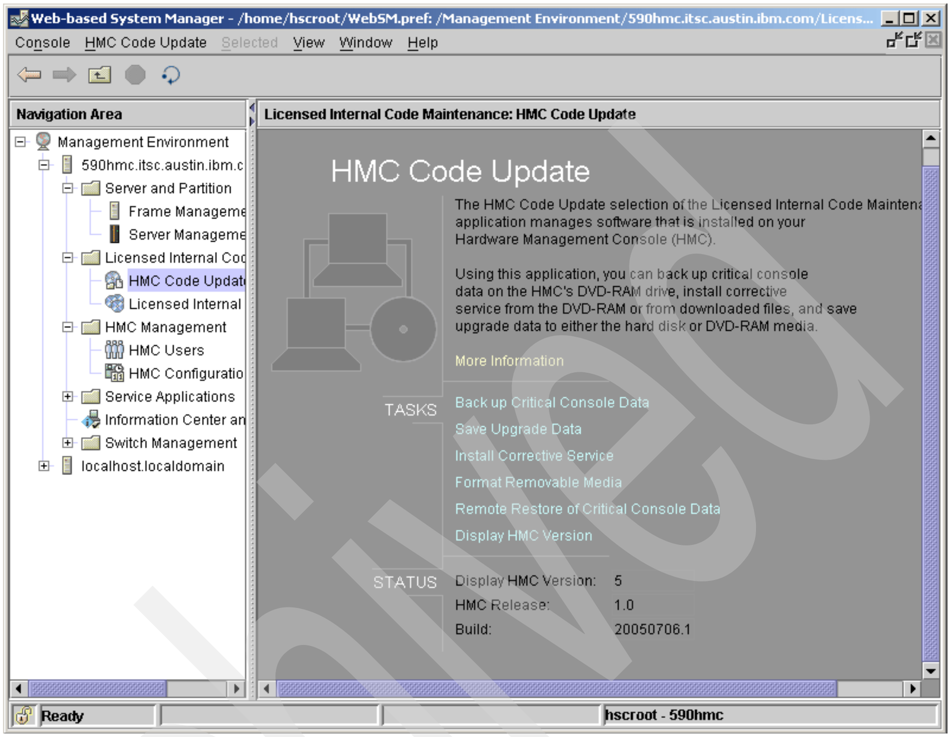
Note: Use the archived data only in conjunction with a reinstallation of the HMC from the product CD-ROMs, for example, after a disk change of the HMC or if the HMC is not responding.

To back up the HMC, perform the following steps:

1. In the Navigation area, click the **Licensed Internal Code Maintenance** icon.
2. In the Contents area, click the **HMC Code Update** icon.
3. Select **Back up Critical Console Data**.
4. Select an archive option. You can back up to DVD on the HMC, to a remote system mounted to the HMC file system (such as NFS), or to a remote site through FTP.
5. Follow the instructions on the panel to back up the data.

Note: The DVD must be formatted in the DVD-RAM format before data can be saved to the DVD.

Figure 4-2 shows the HMC Code Update window.



The HMC Code Update selection of the Licensed Internal Code Maintena application manages software that is installed on your Hardware Management Console (HMC).

Using this application, you can back up critical console data on the HMC's DVD-RAM drive, install corrective service from the DVD-RAM or from downloaded files, and save upgrade data to either the hard disk or DVD-RAM media.

Display HMC Version:	5
HMC Release:	1.0
Build:	20050706.1

Figure 4-2 HMC Code Update window

Restoring critical HMC data

Only restore the HMC backup data in conjunction with a reinstallation of the HMC.

To restore the HMC data, you must be a member of one of the following roles: Super Administrator, Operator, or Service Representative.

Perform the following steps:

1. Shut down and power off the HMC.
2. Power on the HMC console and insert the HMC recovery CD. The HMC powers on from the media and displays the recovery panel.
3. Press F8 to select the **1 - Install/Recover** option.
4. After the installation of the first CD completes, you are prompted to insert the second installation CD into the DVD drive. Press any key to continue. The HMC reboots.

5. After the installation of the second CD completes, select **1 - Install additional software from CD media** from the menu to install the information center from the third CD.
6. After the information center installation, select **1 - Restore Critical Console Data** from the menu to restore data from your backup DVD. To restore from a remote server, select **2**.

Save Upgrade Data task

You can save the current HMC configuration in a special disk partition on the HMC. Only save upgrade data immediately prior to upgrading your HMC software to a new release. This enables you to restore HMC configuration settings after upgrading.

> **Note:** The special disk partition can hold only one level of backup data. Every time you perform this task, previous backup data is overwritten by the latest backup. To proceed with the upgrade after saving upgrade data, you can must place the new HMC recovery CD in the DVD drive and immediately reboot the HMC. Any configuration changes on the HMC after you saved upgrade data will not be saved.

To save the upgrade data, perform the following steps:

1. Expand the Licensed Internal Code Maintenance folder, and then select the HMC application in the Navigation area.
2. Select the **Save Upgrade Data** task in the Content area. An information window opens that prompts you to select the media (hard drive or DVD).
3. Select the appropriate media.
4. Click **Continue**.
5. Click **OK** to confirm and close the information window.

Working with partition profile information

You can back up, restore, initialize, and remove profiles that you have created. These profiles are stored on the HMC hard disk. This section describes each of these options.

Figure 4-3 shows the Profile Data menu.

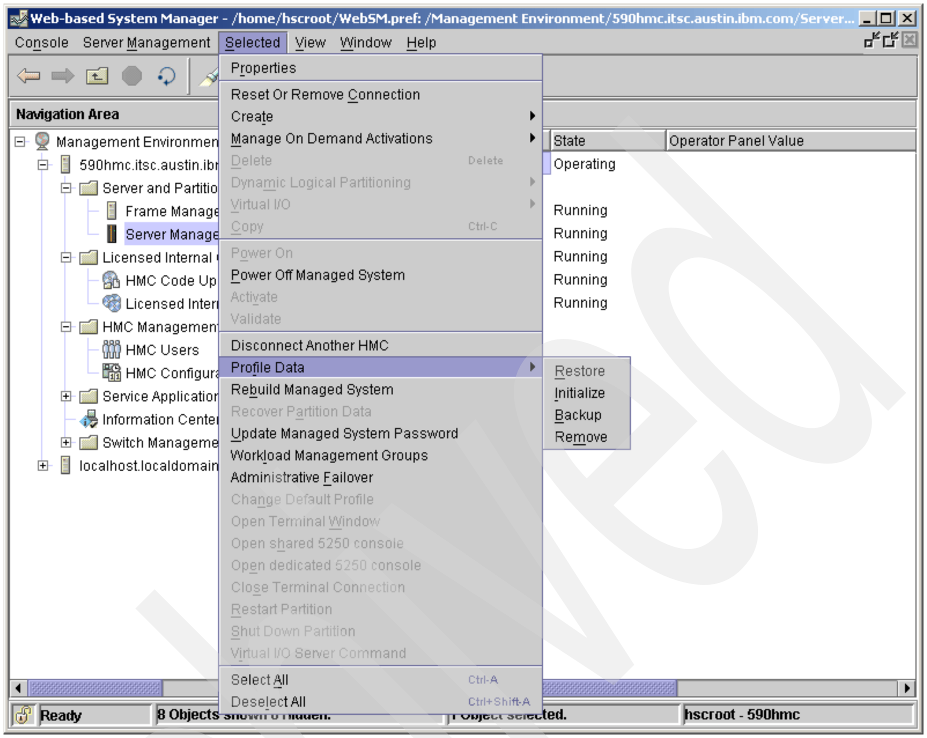


Figure 4-3 Profile Data functions

Backing up partition profile data

To back up partition profile data, you must be a member of one of the following roles: Super Administrator or Service Representative.

To back up partition profile data, perform the following steps:

1. In the Contents area, select the managed system.
2. From the menu, select **Selected** → **Profile Data** → **Backup**.
3. Type the name you want to use for this backup file.
4. Click **OK**.

Initializing profile data

When you initialize profile data, you return the managed system to a state that does not have any logical partitions or profiles. You can perform this task in order to stabilize your managed system if the profile data becomes corrupt.

Important: After you perform this task, any profiles that you created prior to initialization are erased. Use this procedure only under the direction of your service provider.

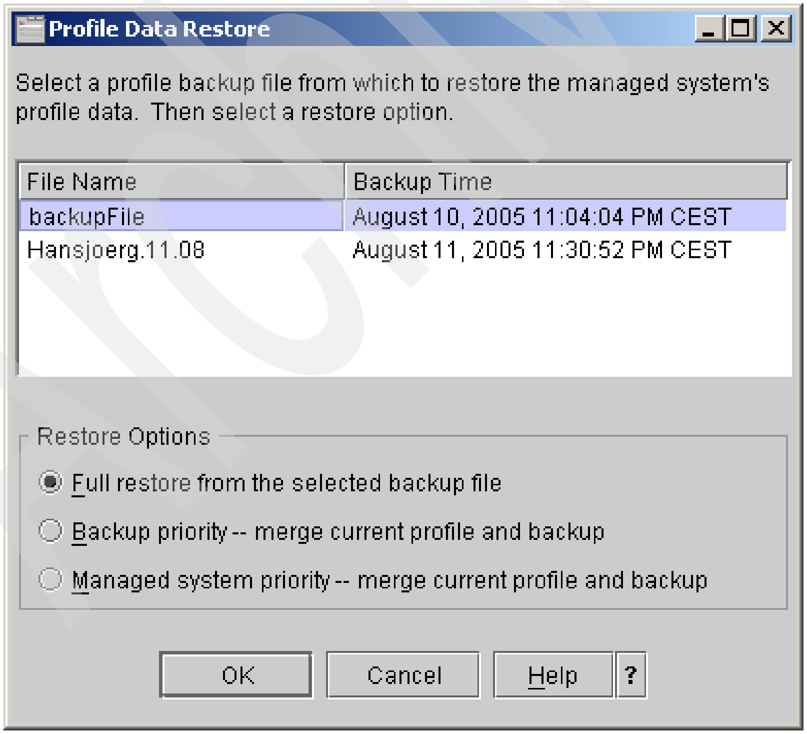
Note: You can initialize profile data only when the managed system is in the Operating or Standby state and all the partitions are in the Not Activated state.

Restoring profile data

Selecting this menu item restores profile data to the HMC from a backup file stored on the HMC hard drive. There are different options to restore the saved data.

Note: This is not a concurrent procedure. When the data is restored, the managed system powers on to Partition Standby.

Figure 4-4 shows the Profile Data Restore window with the options.



Profile Data Restore

Select a profile backup file from which to restore the managed system's profile data. Then select a restore option.

File Name	Backup Time
backupFile	August 10, 2005 11:04:04 PM CEST
Hansjoerg.11.08	August 11, 2005 11:30:52 PM CEST

Restore Options

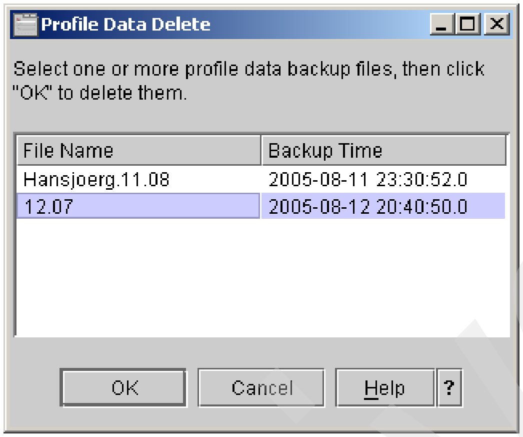
- ● Full restore from the selected backup file
- ○ Backup priority -- merge current profile and backup
- ○ Managed system priority -- merge current profile and backup

Figure 4-4 Profile Data Restore window

Removing profile data

This option is used to remove a stored profile from the HMC hard disk drive.

Figure 4-5 shows the Profile Data Delete window.



Select one or more profile data backup files, then click "OK" to delete them.

File Name	Backup Time
Hansjoerg.11.08	2005-08-11 23:30:52.0
12.07	2005-08-12 20:40:50.0

Figure 4-5 Profile Data Delete window

4.2 Service applications

Remote support enables connectivity to IBM from the HMC. IBM Electronic Service Agent must to be enabled to allow the HMC to connect to IBM to transmit the inventory of the managed system to IBM. Enabling remote support also allows for the reporting of problems to IBM using the HMC. Remote support must be enabled if IBM needs to connect to the HMC for remote servicing. This remote servicing is always initiated from the managed system end rather than the IBM end for security reasons. The following options are available for remote support:

- Client information

  Used to enter the client contact information, such as address, phone numbers, contact person.

- Outbound connectivity settings

  The information required to make a connection to IBM form the HMC for problem reporting and Electronic Service Agent inventory transmissions.

- Inbound connectivity settings

  The information needed for IBM to connect to the HMC for remote service.

- E-mail settings

  This option is used to set a notification by e-mail when the HMC reports a problem to IBM. The user defines what e-mail address will receive the notification.

- Remote support requests
- Remote connections

The Electronic Service Agent application monitors your servers. If the Electronic Service Agent is installed on your HMC, the HMC can monitor all the managed servers. If a problem occurs and the application is enabled, Electronic Service Agent can report the problem to your service and support organization. If your server is partitioned, Electronic Service Agent, together with the Service Focal Point application, reports serviceable events and associated data collected by Service Focal Point to your service and support organization.

4.2.1 Electronic Service Agent

The Electronic Service Agent Gateway maintains the database for all the Electronic Service Agent data and events that are sent to your service and support organization, including any Electronic Service Agent data from other client HMCs in your network. You can enable Electronic Service Agent to perform the following tasks:

- Report problems automatically; service calls are placed without intervention.
- Automatically send service information to your service and support organization.
- Automatically receive error notification either from Service Focal Point or from the operating system running in a Full System Partition profile.
- Support a network environment with a minimum number of telephone lines for modems.

For the client, we can define e-mail notification to the e-mail account.

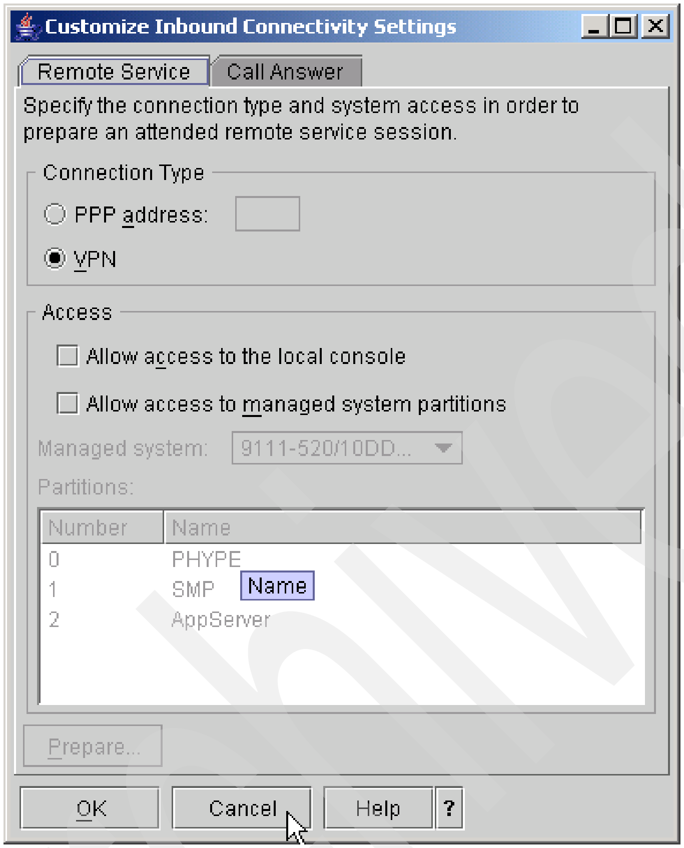
Remote Support Facility

The Remote Support Facility is an application that runs on the HMC and enables the HMC to call out to a service or support facility. The connection between the HMC and the remote facility can be used to:

- Allow automatic problem reporting to your service and support organization
- Allow remote support center personnel to directly access your server in the event of a problem

Ask the client for this information, or fill it out together with the client.

Figure 4-6 shows an example of a VPN connection.



Customize Inbound Connectivity Settings

Specify the connection type and system access in order to prepare an attended remote service session.

Connection Type

○ PPP address:

● VPN

Access

□ Allow access to the local console

□ Allow access to managed system partitions

Managed system: 9111-520/10DD...

Partitions:

Number	Name
0	PHYPE
1	SMP
2	AppServer

Figure 4-6 VPN connection

4.2.2 Service Focal Point

The Service Focal Point (SFP) application is used to help service personnel to diagnose and repair problems on partitioned systems. Service personnel use the HMC as the starting point for all hardware service issues. The HMC gathers various hardware system-management issues at one control point, allowing service personnel to use the Service Focal Point application to determine an appropriate hardware service strategy. Traditional service strategies become more complicated in a logically partitioned environment. Each logical partition runs on its own, unaware that other logical partitions exist on the same system. If a shared resource such as a power supply has an error, the same error might be reported by each partition using the shared resource. The Service Focal Point application enables service personnel to avoid long lists of repetitive call-home information by recognizing that these errors repeat and by filtering them into one serviceable event. To keep the SFP running, we need an Ethernet or virtual Ethernet connection from the HMC to each LPAR.

The following options are available under Service Focal Point:

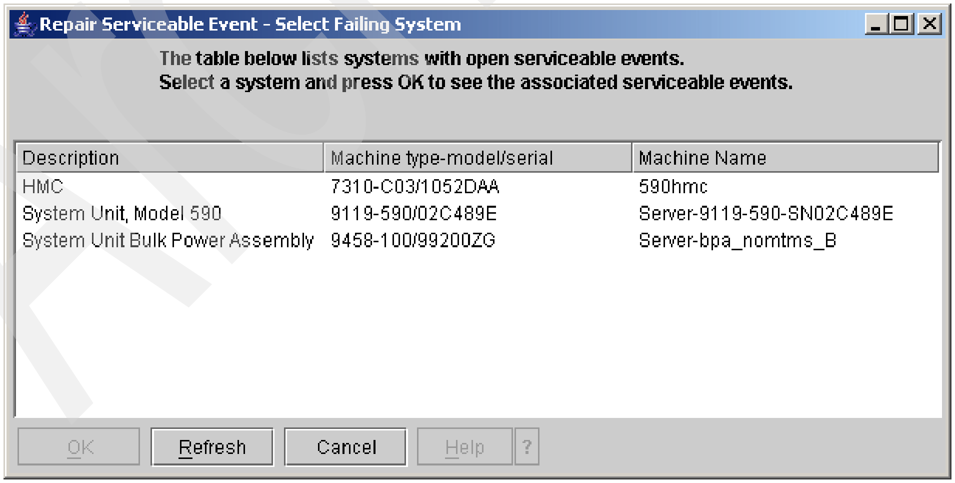
- Repair serviceable event
- Manage serviceable events
- Add/install/remove hardware
- Replace parts
- Service utilities

Repair serviceable event

This option enables the user or the service representative to view a serviceable event and then initiate a repair against that service event. Here, we provide an example of the steps taken to view an event and initiate a repair.

The Service Focal Point is a system infrastructure on the HMC that manages serviceable event information for the system building blocks. It includes resource managers that monitor and record information about different objects in the system. It is designed to filter and correlate events from the resource managers and initiate a call to the service provider when appropriate. It also provides a user interface that enables a user to view the events and perform problem analysis.

Figure 4-7 shows how to select a system.

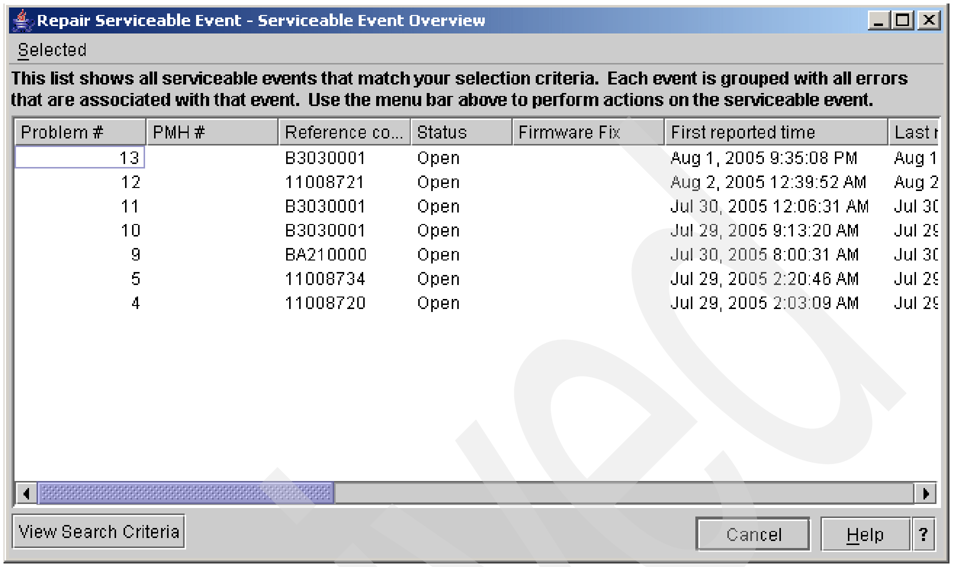


The table below lists systems with open serviceable events.
Select a system and press OK to see the associated serviceable events.

Description	Machine type-model/serial	Machine Name
HMC	7310-C03/1052DAA	590hmc
System Unit, Model 590	9119-590/02C489E	Server-9119-590-SN02C489E
System Unit Bulk Power Assembly	9458-100/99200ZG	Server-bpa_nomtms_B

Figure 4-7 Select a system

Figure 4-8 shows selecting the service events to see more data.

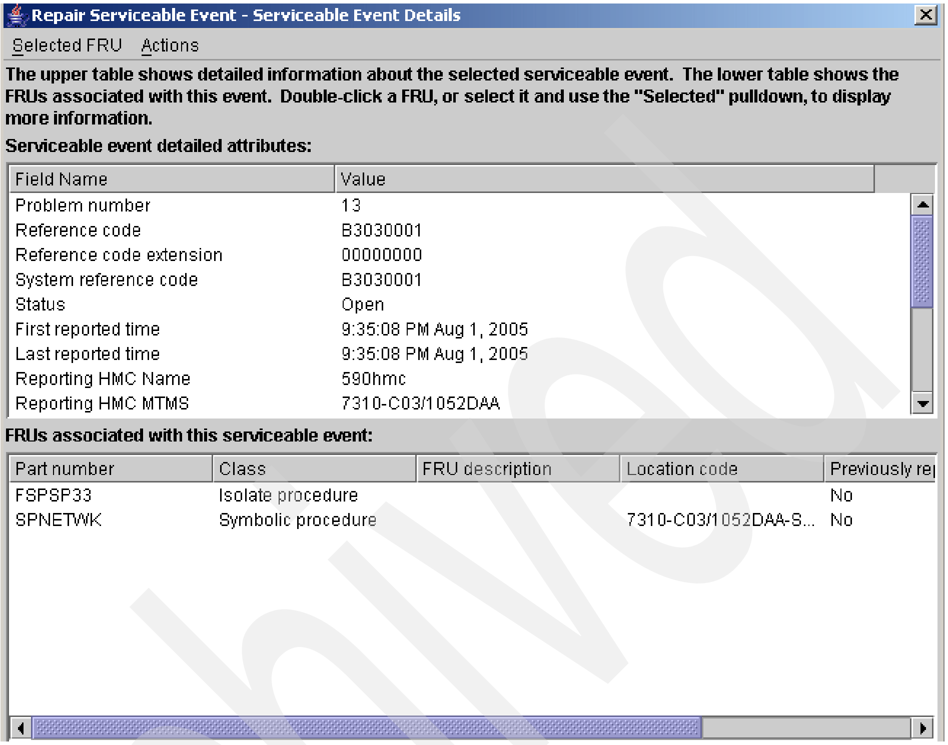


This list shows all serviceable events that match your selection criteria. Each event is grouped with all errors that are associated with that event. Use the menu bar above to perform actions on the serviceable event.

Problem #	PMH #	Reference co...	Status	Firmware Fix	First reported time	Last r
13		B3030001	Open		Aug 1, 2005 9:35:08 PM	Aug 1
12		11008721	Open		Aug 2, 2005 12:39:52 AM	Aug 2
11		B3030001	Open		Jul 30, 2005 12:06:31 AM	Jul 30
10		B3030001	Open		Jul 29, 2005 9:13:20 AM	Jul 29
9		BA210000	Open		Jul 30, 2005 8:00:31 AM	Jul 30
5		11008734	Open		Jul 29, 2005 2:20:46 AM	Jul 29
4		11008720	Open		Jul 29, 2005 2:03:09 AM	Jul 29

Figure 4-8 Manage service events

Figure 4-9 shows details of an event.



The upper table shows detailed information about the selected serviceable event. The lower table shows the FRUs associated with this event. Double-click a FRU, or select it and use the "Selected" pulldown, to display more information.

Serviceable event detailed attributes:

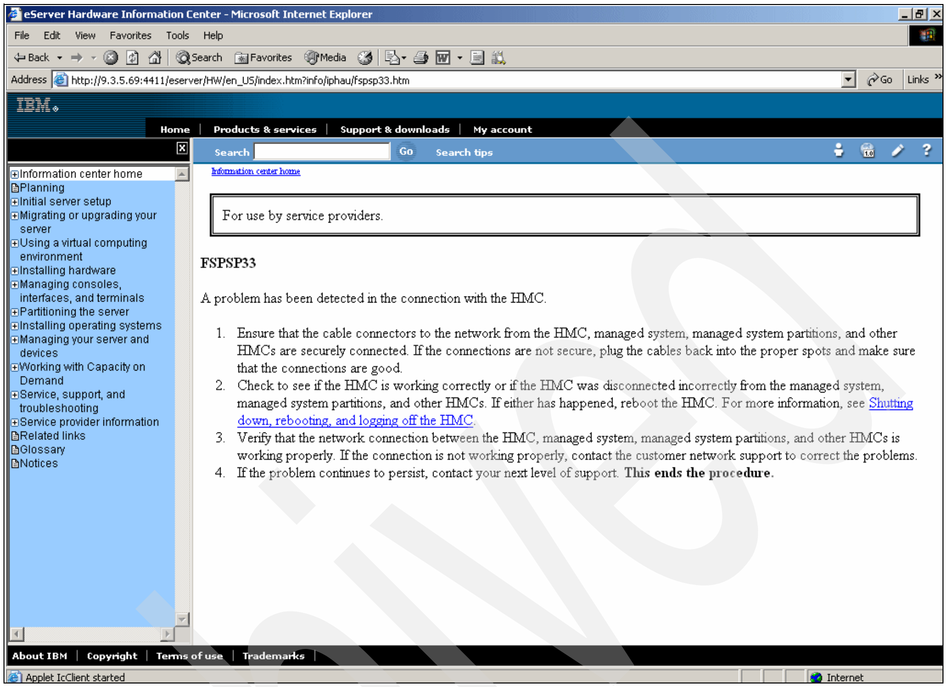
Field Name	Value
Problem number	13
Reference code	B3030001
Reference code extension	00000000
System reference code	B3030001
Status	Open
First reported time	9:35:08 PM Aug 1, 2005
Last reported time	9:35:08 PM Aug 1, 2005
Reporting HMC Name	590hmc
Reporting HMC MTMS	7310-C03/1052DAA

FRUs associated with this serviceable event:

Part number	Class	FRU description	Location code	Previously re
FSPSP33	Isolate procedure			No
SPNETWK	Symbolic procedure		7310-C03/1052DAA-S...	No

Figure 4-9 Detailed view of a service event

Figure 4-10 shows an example from the Information Center for this event.


eServer Hardware Information Center - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://9.3.5.69:4411/eserver/HW/en_US/index.htm?info/iphau/fspsp33.htm
Go Links
IBM.
Home | Products & services | Support & downloads | My account
Search Go Search tips
Information center home
Planning
Initial server setup
Migrating or upgrading your server
Using a virtual computing environment
Installing hardware
Managing consoles, interfaces, and terminals
Partitioning the server
Installing operating systems
Managing your server and devices
Working with Capacity on Demand
Service, support, and troubleshooting
Service provider information
Related links
Glossary
Notices
About IBM | Copyright | Terms of use | Trademarks
Applet IcClient started
Internet

Information center home

For use by service providers.

FSPSP33

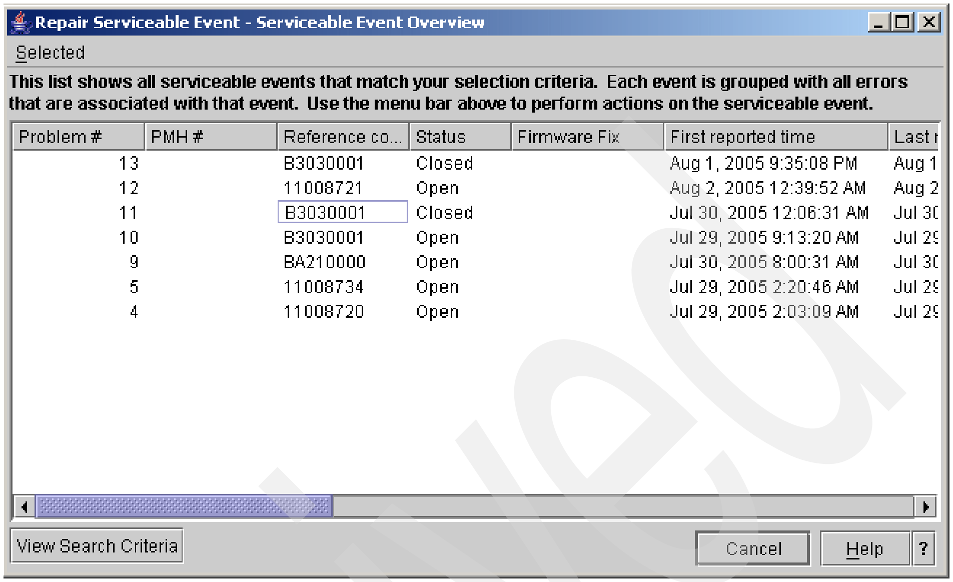
A problem has been detected in the connection with the HMC.

1. Ensure that the cable connectors to the network from the HMC, managed system, managed system partitions, and other HMCs are securely connected. If the connections are not secure, plug the cables back into the proper spots and make sure that the connections are good.
2. Check to see if the HMC is working correctly or if the HMC was disconnected incorrectly from the managed system, managed system partitions, and other HMCs. If either has happened, reboot the HMC. For more information, see Shutting down, rebooting, and logging off the HMC.
3. Verify that the network connection between the HMC, managed system, managed system partitions, and other HMCs is working properly. If the connection is not working properly, contact the customer network support to correct the problems.
4. If the problem continues to persist, contact your next level of support. **This ends the procedure.**

Figure 4-10 Example from Information Center for an event

After following these steps, the event is closed.

Figure 4-11 shows two closed events.



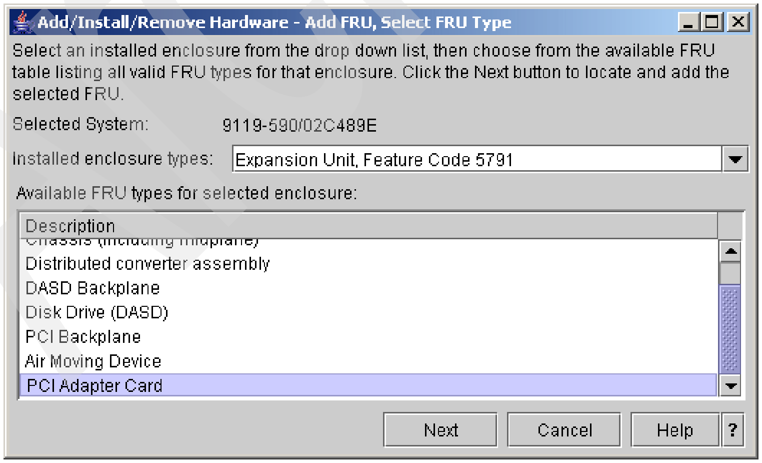
This list shows all serviceable events that match your selection criteria. Each event is grouped with all errors that are associated with that event. Use the menu bar above to perform actions on the serviceable event.

Problem #	PMH #	Reference co...	Status	Firmware Fix	First reported time	Last r
13		B3030001	Closed		Aug 1, 2005 9:35:08 PM	Aug 1
12		11008721	Open		Aug 2, 2005 12:39:52 AM	Aug 2
11		B3030001	Closed		Jul 30, 2005 12:06:31 AM	Jul 3
10		B3030001	Open		Jul 29, 2005 9:13:20 AM	Jul 2
9		BA210000	Open		Jul 30, 2005 8:00:31 AM	Jul 3
5		11008734	Open		Jul 29, 2005 2:20:46 AM	Jul 2
4		11008720	Open		Jul 29, 2005 2:03:09 AM	Jul 2

Figure 4-11 Closed events

Add/install/remove hardware

The Add/Install/Remove Hardware option enables the user or the service representative to add or remove hardware. Figure 4-12 shows the Add/Install/Remove Hardware window.



Select an installed enclosure from the drop down list, then choose from the available FRU table listing all valid FRU types for that enclosure. Click the Next button to locate and add the selected FRU.

Selected System: 9119-590/02C489E

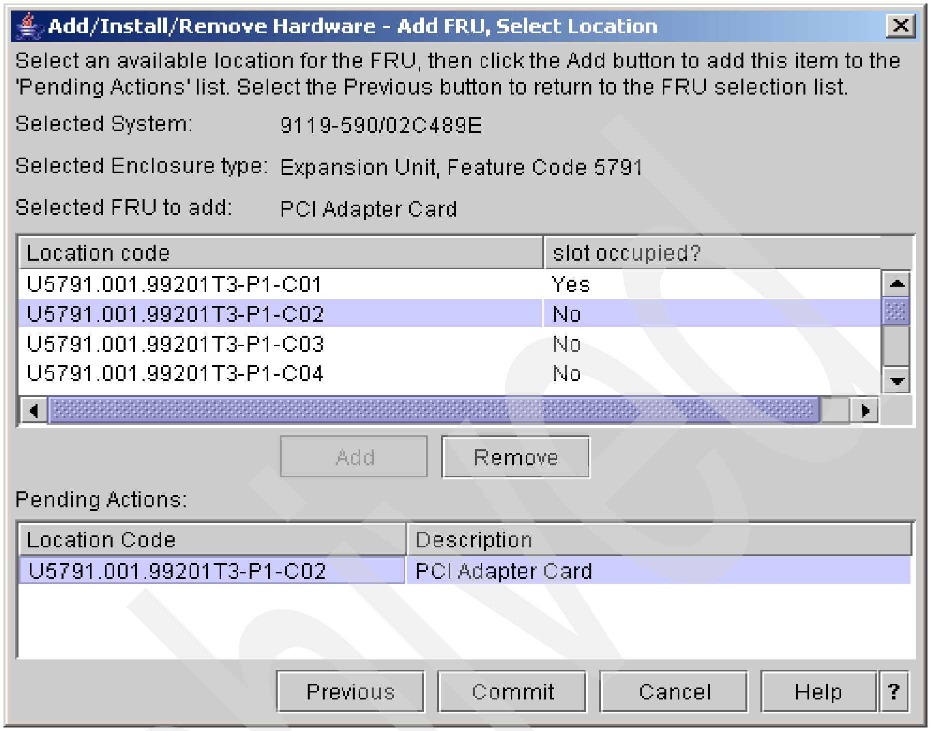
Installed enclosure types: Expansion Unit, Feature Code 5791

Available FRU types for selected enclosure:

Description
Chassis (including midplane)
Distributed converter assembly
DASD Backplane
Disk Drive (DASD)
PCI Backplane
Air Moving Device
PCI Adapter Card

Figure 4-12 Add/Install/Remove Hardware

In the window shown in Figure 4-13, you can select a slot.



Add/Install/Remove Hardware - Add FRU, Select Location

Select an available location for the FRU, then click the Add button to add this item to the 'Pending Actions' list. Select the Previous button to return to the FRU selection list.

Selected System:	9119-590/02C489E
Selected Enclosure type:	Expansion Unit, Feature Code 5791
Selected FRU to add:	PCI Adapter Card

Location code	slot occupied?
U5791.001.99201T3-P1-C01	Yes
U5791.001.99201T3-P1-C02	No
U5791.001.99201T3-P1-C03	No
U5791.001.99201T3-P1-C04	No

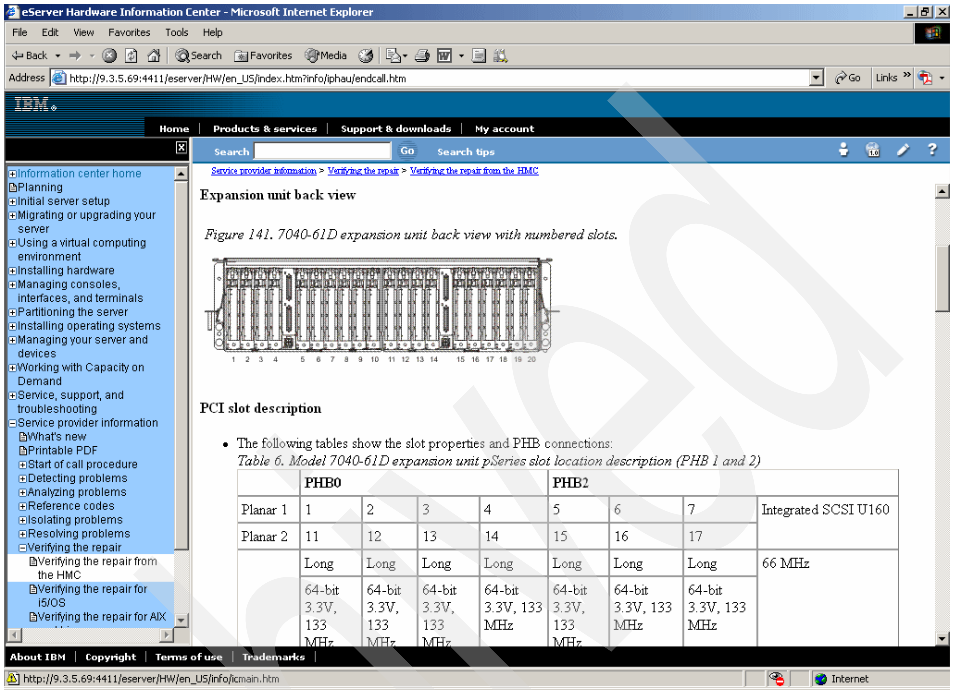
Pending Actions:

Location Code	Description
U5791.001.99201T3-P1-C02	PCI Adapter Card

Figure 4-13 Select the slot

This procedure guides you in the Information Center on the HMC.

Figure 4-14 shows a page from the Information Center. Here you can find details about the adapters and the slots.


eServer Hardware Information Center - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://9.3.5.69:4411/eserver/HW/en_US/index.htm?info/iphau/endcall.htm
Go Links
IBM
Home | Products & services | Support & downloads | My account
Search Go Search tips
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
About IBM | Copyright | Terms of use | Trademarks |
http://9.3.5.69:4411/eserver/HW/en_US/info/icmain.htm
Internet

- Information center home
- Planning
- Initial server setup
- Migrating or upgrading your server
- Using a virtual computing environment
- Installing hardware
- Managing consoles, interfaces, and terminals
- Partitioning the server
- Installing operating systems
- Managing your server and devices
- Working with Capacity on Demand
- Service, support, and troubleshooting
- Service provider information
  - What's new
  - Printable PDF
  - Start of call procedure
  - Detecting problems
  - Analyzing problems
  - Reference codes
  - Isolating problems
  - Resolving problems
  - Verifying the repair
    - Verifying the repair from the HMC
    - Verifying the repair for i5/OS
    - Verifying the repair for AIX

Service provider information > Verifying the repair > Verifying the repair from the HMC

Expansion unit back view

Figure 141. 7040-61D expansion unit back view with numbered slots.

PCI slot description

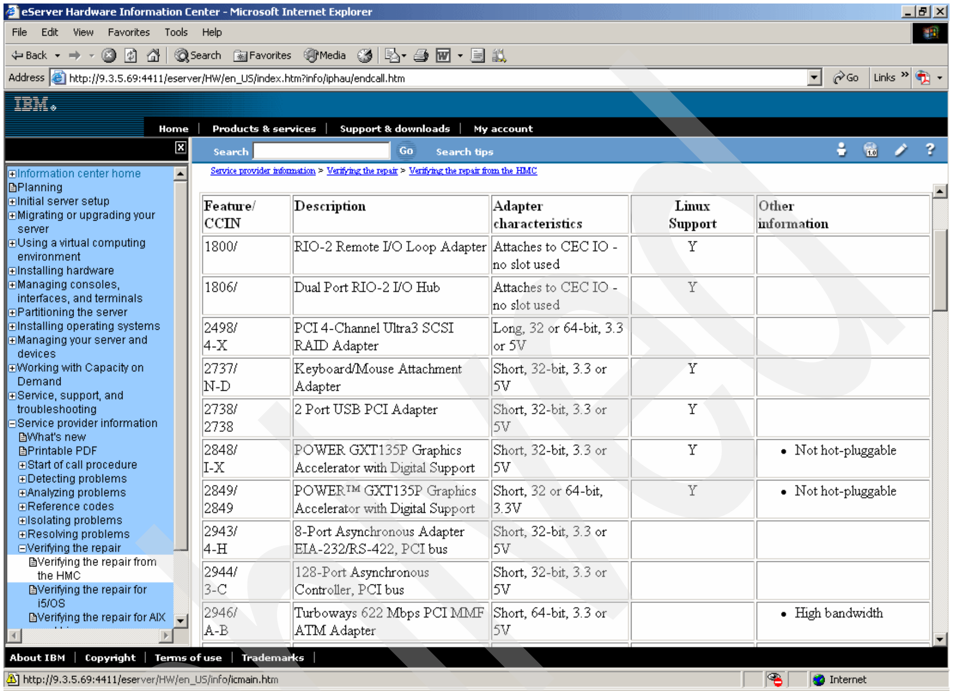
- The following tables show the slot properties and PHB connections:

Table 6. Model 7040-61D expansion unit pSeries slot location description (PHB 1 and 2)

	PHB0				PHB2			
Planar 1	1	2	3	4	5	6	7	Integrated SCSI U160
Planar 2	11	12	13	14	15	16	17	
	Long	Long	Long	Long	Long	Long	Long	66 MHz
	64-bit 3.3V, 133 MHz	64-bit 3.3V, 133 MHz	64-bit 3.3V, 133 MHz	64-bit 3.3V, 133 MHz	64-bit 3.3V, 133 MHz	64-bit 3.3V, 133 MHz	64-bit 3.3V, 133 MHz	

Figure 4-14 Expansion back view with PCI slot description

Figure 4-15 gives you information about Linux support of the adapter and if the adapter is hot-pluggable. All adapters are hot-pluggable except the graphics adapter and the 2-Port Multiprotocol PCI Adapter.



eServer Hardware Information Center - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://9.3.5.69:4411/eserver/HW/en_US/index.htm?info/iphau/endcall.htm
Go Links
IBM.
Home | Products & services | Support & downloads | My account
Search Go Search tips
Information center home
Planning
Initial server setup
Migrating or upgrading your server
Using a virtual computing environment
Installing hardware
Managing consoles, interfaces, and terminals
Partitioning the server
Installing operating systems
Managing your server and devices
Working with Capacity on Demand
Service, support, and troubleshooting
Service provider information
What's new
Printable PDF
Start of call procedure
Detecting problems
Analyzing problems
Reference codes
Isolating problems
Resolving problems
Verifying the repair
Verifying the repair from the HMC
Verifying the repair for i5/OS
Verifying the repair for AIX
Service provider information > Verifying the repair > Verifying the repair from the HMC
About IBM | Copyright | Terms of use | Trademarks
http://9.3.5.69:4411/eserver/HW/en_US/info/icmain.htm
Internet

Feature/ CCIN	Description	Adapter characteristics	Linux Support	Other information
1800/	RIO-2 Remote I/O Loop Adapter	Attaches to CEC IO - no slot used	Y	
1806/	Dual Port RIO-2 I/O Hub	Attaches to CEC IO - no slot used	Y	
2498/ 4-X	PCI 4-Channel Ultra3 SCSI RAID Adapter	Long, 32 or 64-bit, 3.3 or 5V		
2737/ N-D	Keyboard/Mouse Attachment Adapter	Short, 32-bit, 3.3 or 5V	Y	
2738/ 2738	2 Port USB PCI Adapter	Short, 32-bit, 3.3 or 5V	Y	
2848/ I-X	POWER GXT135P Graphics Accelerator with Digital Support	Short, 32-bit, 3.3 or 5V	Y	• Not hot-pluggable
2849/ 2849	POWER™ GXT135P Graphics Accelerator with Digital Support	Short, 32 or 64-bit, 3.3V	Y	• Not hot-pluggable
2943/ 4-H	8-Port Asynchronous Adapter EIA-232/RS-422, PCI bus	Short, 32-bit, 3.3 or 5V		
2944/ 3-C	128-Port Asynchronous Controller, PCI bus	Short, 32-bit, 3.3 or 5V		
2946/ A-B	Turboways 622 Mbps PCI MMF ATM Adapter	Short, 64-bit, 3.3 or 5V		• High bandwidth

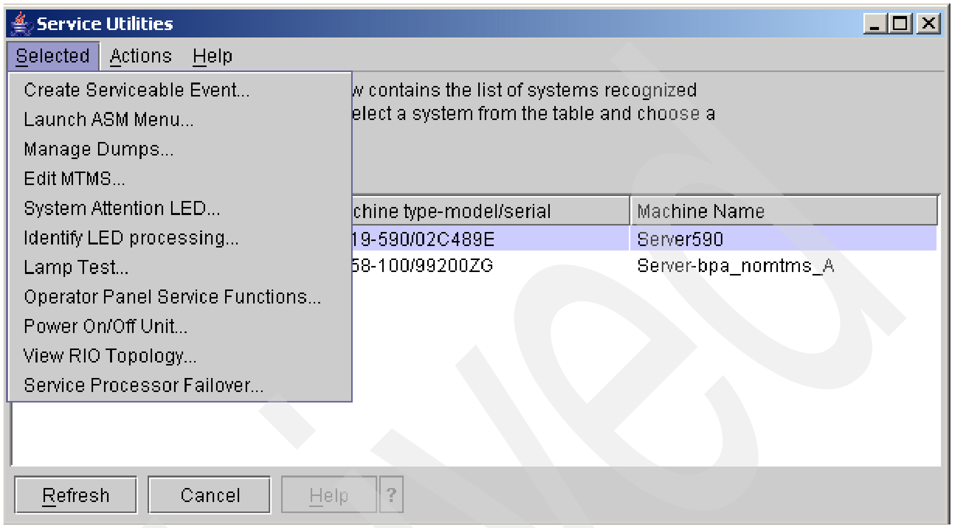
Figure 4-15 More details about the adapter

Replace parts

This option enables the user or service representative to replace parts. The procedure is similar to the Add/Install/Remove Hardware option.

Service utilities

This option is normally used by the service representative. Figure 4-16 shows the options on the Service Utilities window.



w contains the list of systems recognized
elect a system from the table and choose a

chine type-model/serial	Machine Name
19-590/02C489E	Server590
58-100/99200ZG	Server-bpa_nomtms_A

Figure 4-16 Service Utilities window

4.2.3 Licensed Internal Code maintenance

IBM will periodically release firmware updates for the pSeries systems. These updates provide changes to your software, Licensed Internal Code (LIC), or machine code that fix known problems, add new function, and keep your server or Hardware Management Console operating efficiently. For example, you might install fixes for your operating system in the form of a program temporary fix (PTF). Or, you might install a server firmware fix with code changes that are needed to support new hardware or new functions of the existing hardware.

A good fix strategy is an important part of maintaining and managing your server. You should install fixes on a regular basis if you have a dynamic environment that changes frequently. If you have a stable environment, you do not have to install fixes as frequently. However, you should consider installing fixes whenever you make any major software or hardware changes in your environment.

For additional information about Licensed Internal Code maintenance and a example of how to load the code on a p-590/595, go to 2.8, “Firmware” on page 51.

Licensed Internal Code updates

Detailed firmware updating information can be found in the IBM eserver Hardware Information Center.

Firmware (Licensed Internal Code) fixes

This topic describes the following types of firmware (Licensed Internal Code) fixes. For more details and an example how to load the code, see 2.8, “Firmware” on page 51.

Server firmware

Server firmware is the part of the Licensed Internal Code that enables hardware, such as the service processor. Check for available server firmware fixes regularly, and download and install the fixes if necessary. Depending on your service environment, you can download, install, and manage your server firmware fixes using different interfaces and methods, including the HMC or by using functions specific to your operating system. However, if you have a pSeries server that is managed by an HMC, you must use the HMC to install server firmware fixes.

Power subsystem firmware

Power subsystem firmware is the part of the Licensed Internal Code that enables the power subsystem hardware. You must use an HMC to update or upgrade power subsystem firmware fixes.

I/O adapter and device firmware fixes

I/O adapter and device firmware is the part of the Licensed Internal Code that enables hardware, such as Ethernet PCI adapters or disk drives:

- AIX 5L

  If you use an HMC to manage your server, you can use the HMC interface to download and install your I/O adapter and device firmware fixes. If you do not use an HMC to manage your server, you can use the functions specific to your operating system to work with I/O adapter and device firmware fixes.

- i5/OS

  I/O adapter and device firmware PTFs for i5/OS partitions are ordered, packaged, delivered, and installed as part of the Licensed Internal Code using the same processes that apply to i5/OS PTFs. Regardless of whether you use an HMC to manage your server, you use the usual i5/OS PTF installation functions on each logical partition to download and install the I/O adapter and device firmware fixes.

Notes: For all models except model 57x and 59x servers, if you use an HMC to manage your system and you are setting up the server for the first time or upgrading to a new server firmware release, we recommend that you install the HMC fixes before you install server firmware fixes so that the HMC can handle any fixes or new function updates that you apply to the server.

For model 57x and 59x servers, you must install HMC fixes before you install server or power subsystem fixes so that the HMC can handle any fixes or new function updates that you apply to the server.

4.3 LPAR Validation Tool

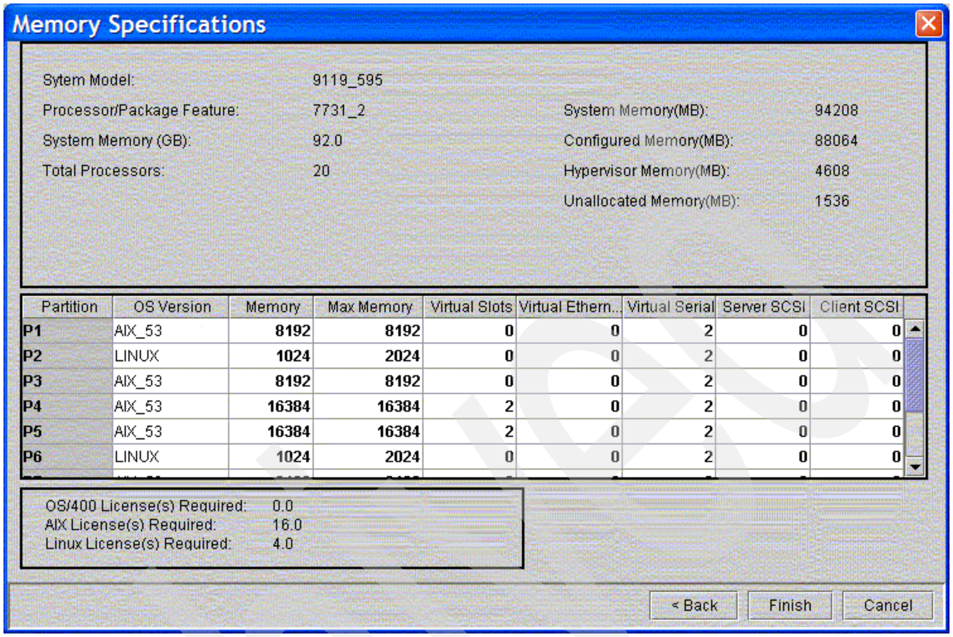
When configuring dynamic or virtual partitions on eServer p5 systems, you can use the LPAR Validation Tool (LVT) to verify system resource requirements. With the LVT, you can customize the partition design by selecting PCI slots for given adapters, specific drives to selected bays, and much more. The LVT provides a useful report that can complement the organization and validation of features required for configuration of a complex partition solution. The LVT supports IBM eServer p5 and eServer i5 servers, iSeries™, and OpenPower systems. A proficient knowledge of LPAR design requirements, limitations, and best practices facilitates the use of this tool.

The LVT tool provides the following functions:

- Supports partitions running AIX 5L Version 5.2 and Version 5.3, Linux, and i5/OS.
- Validates of dynamic LPAR design.
- Validates of virtual partition design, including Virtual I/O Server and virtual clients.
- Calculates unallocated memory and shared processor pool.
- Calculates Hypervisor memory requirements.
- Calculates number of operating system licenses needed to support partition design.
- Validates number of virtual slots required for partitions.

Important: We recommend the use of the LVT to calculate Hypervisor requirements to determine memory resources required for all partitioned and non-partitioned servers.

Figure 4-17 shows the calculated Hypervisor memory requirements based on sample partition requirements.



Sytem Model:	9119_595		
Processor/Package Feature:	7731_2	System Memory(MB):	94208
System Memory (GB):	92.0	Configured Memory(MB):	88064
Total Processors:	20	Hypervisor Memory(MB):	4608
		Unallocated Memory(MB):	1536

Partition	OS Version	Memory	Max Memory	Virtual Slots	Virtual Ethern...	Virtual Serial	Server SCSI	Client SCSI
P1	AIX_53	8192	8192	0	0	2	0	0
P2	LINUX	1024	2024	0	0	2	0	0
P3	AIX_53	8192	8192	0	0	2	0	0
P4	AIX_53	16384	16384	2	0	2	0	0
P5	AIX_53	16384	16384	2	0	2	0	0
P6	LINUX	1024	2024	0	0	2	0	0

OS/400 License(s) Required:	0.0
AIX License(s) Required:	16.0
Linux License(s) Required:	4.0

Figure 4-17 LVT window showing Hypervisor requirements

The LVT is a stand-alone Java™ application that runs on a Microsoft Windows® 95 or later with 128 MB minimum of free memory.

For download and installation information, including the *User's Guide*, visit:

`http://www.ibm.com/servers/eserver/iseries/lpar/systemdesign.htm`

5

Ongoing support

This chapter discusses the ongoing support issues for the IBM @server pSeries Enterprise systems. It includes an introduction to the system maintenance possibilities and introduces features of the AIX 5L operating system and hardware performance management.

This chapter discusses the following topics:

- Perform preventative hardware maintenance
- Hardware performance
- AIX 5L performance monitoring
- Manage system software updates
- Basic problem determination
- Logical volume management
- Planning and performing backup and recovery

5.1 Perform preventative hardware maintenance

The next section describes key preventative maintenance steps.

The Electronic Service Agent can report all critical errors to an IBM service representative. When the Electronic Service Agent is not used, it is necessary to check in the HMC for the serviceable events in the Service Focal Point menu and also the AIX error log. For more information about Service Focal Points, see 4.2.2, “Service Focal Point” on page 122.

Especially in a dirty environment, it is necessary to check from time to time the air filter inside the front door of the p5-590 and p5-595 systems.

Dirty filters reduce the air circulation and the system can overheat. Dust accumulation inside servers might require cleaning to prevent errors.

5.2 Hardware performance

The performance of pSeries systems is directly affected by the following factors:

- Memory bandwidth
- I/O bandwidth

Each of these factors can be properly configured for optimal performance, depending on your workload. This section discusses the methods and techniques for achieving optimal performance.

5.2.1 Memory bandwidth

For optimum performance, install memory based on a specific plugging order based on the server.

All memory in a node must be of the same size and type. There are two types of memory in some servers, DDR1 and DDR2.

5.2.2 I/O bandwidth

To maximize I/O bandwidth for a partition, it is important to consider the following guidelines:

- Avoid I/O drawer contention.

  Keep partitions in separate I/O planars as much as possible. If possible, allocate an entire I/O planar to partitions to eliminate I/O contention. When I/O planars are shared, plan to locate high-bandwidth devices on separate PCI host bridges. Use the remaining PCI slots for low bandwidth devices where possible.

- Maximize data bandwidth.

  Use multiple I/O planars and multiple I/O drawers to increase the total bandwidth available to a partition.

- Use dedicated adapters.

  The use of physical adapters will provide more consistent application performance compared to the use of virtual adapters.

5.3 AIX 5L performance monitoring

The performance analysis tools in AIX 5L provides a number of tools designed to monitor and analyze operating and resource performance. The following list shows the most commonly used performance monitoring commands:

truss Traces a process's system calls, received signals, and incurred machine faults.

alstat Shows alignment exception statistics.

iostat Reports central processing unit (CPU) statistics and I/O statistics for the entire system, adapters, TTY devices, disks, and CD-ROMs.

vmstat Reports virtual memory statistics.

sar Collects, reports, or saves system activity information.

prof Displays object file profile data.

tprof Reports CPU usage.

gprof Displays call graph profile data.

emstat Shows emulation exception statistics.

filemon Monitors the performance of the file system, and reports the I/O activity on behalf of logical files, virtual memory segments, logical volumes, and physical volumes.

fileplace Displays the placement of file blocks within logical or physical volumes.

netpmon Monitors activity and reports statistics on network I/O and network-related CPU usage.

pprof Reports CPU usage of all kernel threads over a period of time.

rmss Simulates a system with various sizes of memory for performance testing of applications.

svmon Captures and analyzes a snapshot of virtual memory.

topas Reports selected local system statistics.

lparstat Reports LPAR-related information and utilization statistics.

mpstat Reports performance statistics for all logical CPUs in the system.

5.3.1 CPU analysis and tuning

When investigating a performance problem, we usually start by monitoring the CPU utilization statistics. It is important continuously observe system performance, because you have to compare the loaded system data with normal usage data.

Generally, CPU is one of the fastest components of the system, and if CPU utilization keeps the CPU 100% busy, this also affects system-wide performance. If you discover that the system keeps the CPU 100% busy, you need to investigate the process that causes this. AIX 5L provides many trace and profiling tools for the system and processes.

5.3.2 Performance considerations for POWER5-based systems

POWER5 provides new and improved functions for more granular and flexible partitioning. It contain new technologies, such as Micro-Partitioning and simultaneous multithreading (SMT).

Micro-Partitioning provides the ability to share a single processor between multiple partitions. These partitions are called shared processor partitions. POWER5-based systems continue to support partitions with dedicated processors. These partitions are called dedicated partitions. Dedicated partitions do not share a single physical processor with other partitions.

In a shared-partition environment, the POWER Hypervisor schedules and distributes processor entitlement to shared partitions from a set of physical processors. This physical processor set is called a shared processor pool. Processor entitlement is distributed with each turn of the Hypervisor's dispatch wheel, and each partition consumes or cedes the given processor entitlement. Figure 5-1 shows a sample of a dedicated partition and a micro-partition on POWER5-based server.

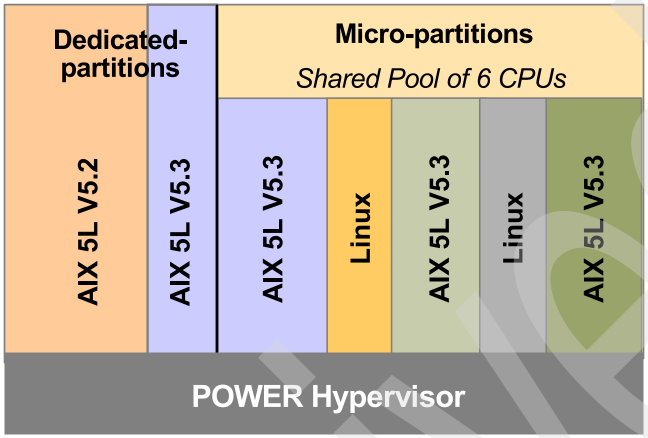


Figure 5-1 LPARs configuration on POWER5-based server

Some monitoring and tuning tools are new in AIX 5L Version 5.3.

- Monitoring tools:
  - **lparstat** (new command in AIX 5L Version 5.3)
  - **mpstat** (new command in AIX 5L Version 5.3)
  - **procmon** (new tool in AIX 5L Version 5.3)
- Tuning tool:
  - **smtctl** (new command in AIX 5L Version 5.3)

5.3.3 Memory analysis and tuning

There are many ways to investigate different parameters and settings, but combining several tools and commands can give you the best overall picture of performance. These commands have many uses. In this section, we only discuss how they can be used to monitor memory. We show how these commands can be used to gauge how the memory of the system is performing at any given moment:

- Memory monitoring:
  - The **ps** command
  - The **sar** command

  - The **svmon** command
  - The **topas** monitoring tool
  - The **vmstat** command
- Memory tuning:
  - The **vmo** command

The svmon command

The **svmon** command is an analysis tool for virtual memory. It captures the current state of memory, including real, virtual, and paging space memory. The **svmon** command invokes the **svmon_back** command. Both are located in /usr/lib/perf, and both part of the perfagent.tools fileset.

Useful combinations of the svmon command include:

- **svmon or svmon -G**
- **svmon -P**
- **svmon -C**
- **svmon -i**

With the **svmon** command, you can display memory usage statistics for processes using the **-P** flag and specifying the process ID (PID). If no PID is supplied, it provides statistics for all active processes. Example 5-1 shows the output of the **svmon -P** command.

Example 5-1 Example using the svmon -P command

```
[p630n04][/]> svmon -P |grep -p Pid
-------------------------------------------------------------------------------
Pid   Command          Inuse      Pin     Pgsp  Virtual 64-bit Mthrd LPage
68532 java             80615     5485        0    29922      N     Y     N
Pid   Command          Inuse      Pin     Pgsp  Virtual 64-bit Mthrd LPage
29290 tnameserv       25022     5471        0    17630      N     Y     N
Pid   Command          Inuse      Pin     Pgsp  Virtual 64-bit Mthrd LPage
15510 hagsd            18305     5487        0    15091      N     N     N
```

Process ID 68532 is using 80615 pages of real memory and no paging space.

The topas command

The **topas** command is a performance monitoring tool that is ideal for broad spectrum performance analysis.

The **topas** command requires that the perfagent.tools fileset is installed on the system. The **topas** command resides in /usr/bin and is part of the bos.perf.tools fileset that is obtained from the AIX 5L base installable media.

Useful combinations of the **topas** command include:

- **topas**
- **topas -i**

The **topas** command is a very useful and common command for CPU performance analyses. It is used to display statistics about the activity on the local system. The **topas** command reports the various kinds of statistics, such as CPU utilization, CPU events and queues, process lists, memory and paging statistics, disk and network performance, and NFS statistics. The **topas** command resides in /usr/bin and is part of the bos.perf.tools fileset, which is installable from the AIX 5L base installation media.

Default output

Starting with AIX 5L Version 5.3, if the **topas** command runs on a shared partition, the following two new values are reported for the CPU utilization. If the **topas** command runs on a dedicated partition, these values are not displayed.

Physc Number of physical processors granted to the partition

%Entc Percentage of entitled capacity granted to the partition

The **topas** command shows following information in default mode:

- System host name
- Current date
- Refresh interval
- CPU utilization
- CPU events and queues
- Process lists
- Memory and paging statistics
- Disk and network performance
- Workload Manger performance (displayed only when Workload Manager is used)
- NFS statistics

Example 5-2 on page 142 shows the standard **topas** command and its output. In this example, it runs on a shared partition. If you run the **topas** command without flags, the output is refreshed every two seconds.

Example 5-2 The topas command

```
Topas Monitor for host:    vio_client2          EVENTS/QUEUES    FILE/TTY
Thu Aug 11 15:09:59 2005   Interval:  2         Cswitch     158  Readch        0
                                                Syscall      57  Writech     115
Kernel    0.5   |#                           |  Reads         0  Rawin         0
User      0.1   |#                           |  Writes        0  Ttyout      115
Wait      0.0   |                            |  Forks         0  Igets         0
Idle     99.4   |############################|  Execs         0  Namei         0
Physc =  0.00                     %Entc=   1.0  Runqueue    0.0  Dirblk        0
                                                Waitqueue   0.0
Network  KBPS   I-Pack  O-Pack   KB-In  KB-Out
en0       0.2      0.5     0.5     0.0     0.1  PAGING           MEMORY
lo0       0.0      0.0     0.0     0.0     0.0  Faults        0  Real,MB     512
                                                Steals        0  % Comp     49.0
Disk    Busy%     KBPS     TPS KB-Read KB-Writ  PgspIn        0  % Noncomp  45.2
hdisk0    0.0      0.0     0.0     0.0     0.0  PgspOut       0  % Client   48.5
                                                PageIn        0
Name            PID  CPU%  PgSp Owner           PageOut       0  PAGING SPACE
topas        294994   0.1   1.4 root            Sios          0  Size,MB     512
syncd         98426   0.0   0.5 root                             % Used     10.9
getty        241866   0.0   0.4 root            NFS (calls/sec)  % Free     89.0
gil           57372   0.0   0.1 root            ServerV2       0
xmgc          45078   0.0   0.0 root            ClientV2       0   Press:
rgsr         127052   0.0   0.0 root            ServerV3       0   "h" for help
rpc.lock     172192   0.0   0.2 root            ClientV3       0   "q" to quit
nmon_aix     229540   0.0   0.8 root
pilegc        40980   0.0   0.1 root
rmcd         217196   0.0   2.5 root
```

The new **-L** flag has been added to the **topas** command to display the logical partition. In this mode, the result of the **topas** command is similar to the **mpstat** command. Example 5-3 shows a sample of the **topas** command with the **-L** flag.

Example 5-3 The topas command with the -L flag

```
Interval:    2    Logical Partition: VIO_client2       Thu Aug 11 15:13:19 2005
Psize:   6                    Shared SMT  ON           Online Memory:   512.0
Ent: 0.50                      Mode: Capped            Online Logical CPUs:  2
Partition CPU Utilization                              Online Virtual CPUs:  1
%usr %sys %wait %idle physc %entc %lbusy   app   vcsw phint %hypv   hcalls
   0    0     0   100   0.0   0.80  0.00  5.98    334     0   0.0        0
===============================================================================
LCPU  minpf majpf  intr   csw icsw runq lpa scalls usr sys _wt idl   pc   lcsw
          0     0     0     0    0    0   0      0   0   0   0   0 0.00      0
          0     0     0     0    0    0   0      0   0   0   0   0 0.00      0
Cpu2      0     0    30   135   66    0 100     24   6  60   0  34 0.00    192
Cpu3      0     0   167     0    0    0   0      0   0  26   0  74 0.00    142
```

Paging statistics

There are two parts of the paging statistics reported by the **topas** command. The first part is total paging statistics. This simply reports the total amount of paging available on the system and the percentages free and used. The second part provides a breakdown of the paging activity. The reported items and their meanings are as follows:

Faults Reports the number of faults.

Steals Reports the number of 4 KB pages of memory stolen by the Virtual Memory Manager per second.

PgspIn Reports the number of 4 KB pages read in from the paging space per second.

PgspOut Reports the number of 4 KB pages written to the paging space per second.

PageIn Reports the number of 4 KB pages read per second.

PageOut Reports the number of 4 KB pages written per second.

Sios Reports the number of I/O requests per second issued by the Virtual Memory Manager.

Memory statistics

The memory statistics are as follows:

Real Shows the actual physical memory of the system in megabytes.

%Comp Reports real memory allocated to computational pages.

%Noncomp Reports real memory allocated to non-computational pages.

%Client Reports on the amount of memory that is currently used to cache remotely mounted files.

The vmstat command

The **vmstat** command is useful for reporting statistics about virtual memory. The **vmstat** command is located in /usr/bin, is part of the bos.acct fileset, and is installable from the AIX 5L base installation media.

The **vmstat** command summarizes the total active virtual memory used by all of the processes in the system, as well as the number of real memory page frames on the free list. Active virtual memory is defined as the number of virtual memory working segment pages that have been touched. This number can be larger than the number of real page frames in the machine, because some of the active virtual memory pages might have been written out to paging space.

Useful combinations of the **vmstat** command include:

- **vmstat**
- **vmstat Interval Count**
- **vmstat -v**

The **vmstat** command provides data about virtual memory activity to standard output. The first line of data is an average since the last system reboot.

Example 5-4 shows a sample of the **vmstat** command.

Example 5-4 The vmstat command

```
#vmstat 1 5

System configuration: lcpu=2 mem=512MB ent=0.50

kthr    memory              page              faults              cpu
----- ----------- ------------------------ ------------ -----------------------
 r  b   avm   fre  re  pi  po  fr   sr  cy  in   sy  cs us sy id wa    pc    ec
 0  0 70722 10604   0   0   0   0    0   0   2   58 145  0  1 99  0  0.01   1.4
 0  0 70722 10604   0   0   0   0    0   0   6   14 142  0  1 99  0  0.00   1.0
 0  0 70722 10604   0   0   0   0    0   0   1   29 149  0  0 99  0  0.00   0.9
 0  0 70722 10604   0   0   0   0    0   0   9  106 174  0  1 99  0  0.01   1.3
 1  0 70722 10604   0   0   0   0    0   0   2    9 147  0  1 99  0  0.01   1.1
```

The lparstat command

The **lparstat** command is a performance monitoring tool that provides a report about LPAR-related information and utilization statistics. This command provides a display of current LPAR-related parameters, POWER Hypervisor information, and utilization statistics for the LPAR.

The **lparstat** command has the following three modes.

Monitoring mode (default)

The **lparstat** command with no options generates a single report containing utilization statistics related to the LPAR since boot time. The utilization statistics provide the following information:

%user Shows the percentage of the entitled processing capacity used while executing at the user (or application) level.

%sys Shows the percentage of the entitled processing capacity used while executing at the system (or kernel) level.

%idle Shows the percentage of the entitled processing capacity unused while the partition was idle and did not have any outstanding disk I/O requests.

%wait Shows the percentage of the entitled processing capacity unused while the partition was idle and had outstanding disk I/O requests.

For the dedicated partitions, the entitled processing capacity is the number of physical processors.

The following statistics are displayed only on the shared partition:

physc Shows the number of physical processors consumed.

%entc Shows the percentage of the entitled capacity consumed.

lbusy Shows the percentage of logical processors utilization while executing at the user and system level.

app Shows the available physical processors in the shared pool.

phint Shows the number of phantom (targeted to another shared partition in this pool) interruptions received.

Example 5-5 shows a sample of the **`lparstat`** command showing a utilization statistics report.

Example 5-5 Displaying the utilization statistics with the lparstat command

```
#lparstat 1 5

System configuration: type=Shared mode=Capped smt=On lcpu=2 mem=512 psize=6 ent=0.50

%user  %sys  %wait  %idle physc %entc  lbusy   app  vcsw phint
----- ----  -----  ----- ----- ----- ------   ---  ---- -----
  0.0   0.3    0.0   99.6  0.00   0.9    0.0  5.99   978     0
  0.0   0.4    0.0   99.6  0.00   0.9    0.0  5.99   992     0
  0.0   0.3    0.0   99.7  0.00   0.8    0.0  5.99   976     0
  0.0   0.3    0.0   99.7  0.00   0.7    0.0  5.99   868     0
  0.0   0.2    0.0   99.8  0.00   0.6    0.0  5.99   660     0
```

Information mode

The **`lparstat`** command with the **`-i`** flag displays static LPAR configuration. Example 5-6 on page 146 shows a sample of a static LPAR configuration report.

Example 5-6 Displaying the static LPAR configuration report

```
#lparstat -i

Node Name                                  : vio_client2
Partition Name                            : VIO_client2
Partition Number                           : 1
Type                                       : Shared-SMT
Mode                                       : Capped
Entitled Capacity                          : 0.50
Partition Group-ID                         : 32769
Shared Pool ID                             : 0
Online Virtual CPUs                        : 1
Maximum Virtual CPUs                       : 32
Minimum Virtual CPUs                       : 1
Online Memory                              : 512 MB
Maximum Memory                             : 1024 MB
Minimum Memory                             : 128 MB
Variable Capacity Weight                   : 0
Minimum Capacity                           : 0.10
Maximum Capacity                           : 2.00
Capacity Increment                         : 0.01
Maximum Physical CPUs in system            : 16
Active Physical CPUs in system             : 8
Active CPUs in Pool                        : 6
Unallocated Capacity                       : 0.00
Physical CPU Percentage                    : 50.00%
Unallocated Weight                         : 0
```

Hypervisor mode

The **lparstat** command with the **-H** flag provides detailed Hypervisor information. This option displays the statistics for each of the Hypervisor calls.

5.3.4 The nmon tool

The AIX community provides a tool called **nmon**. It is a no-charge performance monitoring tool for AIX 5L and Linux, available at:

http://www.ibm.com/developerworks/eserver/articles/analyze_aix/index.html

The **nmon** command gives you information on one screen and can save data to a comma-separated value (.csv) file for later analysis. It is helpful in presenting all the important performance tuning information on one screen and dynamically updating it.

Example 5-7 shows the default output of the **nmon** command.

Example 5-7   The nmon command output

```
#nmon

+------------------------------------------------------------------------------------------+
 ¦  ------------------------------                                                          ¦
 ¦  N    N  M    M   OOOO   N    N   For online help type: h                                ¦
 ¦  NN   N  MM  MM  O    O  NN   N   For command line option help:                          ¦
 ¦  N N  N  M MM M  O    O  N N  N      quick-hint  nmon -?                                 ¦
 ¦  N  N N  M    M  O    O  N  N N    full-details  nmon -h                                 ¦
 ¦  N   NN  M    M  O    O  N   NN   To start nmon the same way every time?                 ¦
 ¦  N    N  M    M   OOOO   N    N    set NMON ksh variable, for example:                   ¦
 ¦  ------------------------------    export NMON=cmt                                       ¦
 ¦       Version v10p for AIX53                                                             ¦
 ¦                               2 - CPUs currently                                         ¦
 ¦                               2 - CPUs configured                                        ¦
 ¦                            1656 - MHz CPU clock rate                                     ¦
 ¦                  PowerPC_POWER5 - Processor                                              ¦
 ¦                          64 bit - Hardware                                               ¦
 ¦                          64 bit - Kernel                                                 ¦
 ¦                         Dynamic - Logical Partition                                      ¦
 ¦                        5.3.0.30 - AIX Kernel Version                                     ¦
 ¦                     vio_client2 - Hostname                                               ¦
 +------------------------------------------------------------------------------------------+
```

Example 5-8 shows the output of **nmon** command with the `m` and **`c`** options showing memory and CPU usage.

Example 5-8   The nmon output for memory and CPU usage

```
--nmon-v10p---a=disk-Adapters----Host=vio_client2----Refresh=2 secs---14:54.24----------------
+-CPU-Utilisation-Small-View------------EntitledCPU=  0.50 UsedCPU=  0.005-------------------+
¦Logical                    0----------25-----------50----------75----------100              ¦
¦CPU User%  Sys% Wait% Idle%|           |            |           |            |              ¦
¦  2   0.0   0.0   0.0 100.0|  >                                              |              ¦
¦  3   0.0   0.0   0.0 100.0|>                                                |              ¦
¦Logical/Physical Averages  +-----------|------------|-----------|------------+              ¦
¦Log   0.0   0.0   0.0 100.0|                                                 |              ¦
¦Phy   8.3  54.1   0.0  37.6|UUUUsssssssssssssssssssssssssss                  |              ¦
¦Entitlement Used=  1.1%    +-----------|------------|-----------|------------+              ¦
+--------------------------------------------------------------------------------------------+
+-Memory-Use---------------------Paging------------------------Stats-------------------------+
¦          Physical PagingSpace         pages/sec  In     Out  FileSystemCache               ¦
¦% Used       95.1%      10.3%   to Paging Space   0.0    0.0  (numperm)  45.8%              ¦
¦% Free        4.9%      89.7%   to File System    0.0    0.0  Process    15.5%              ¦
¦MB Used     486.9MB     52.9MB  Page Scans        0.0         System     33.8%              ¦
¦MB Free      25.1MB    459.1MB  Page Cycles       0.0         Free        4.9%              ¦
¦Total(MB)   512.0MB    512.0MB  Page Steals       0.0                   ------              ¦
¦                                Page Faults       0.0         Total     100.0%              ¦
¦Min/Maxperm      92MB( 18%)  368MB( 72%) note: % of memory                                  ¦
¦Min/Maxfree     960   1088      Total  Virtual    1.0GB       User       55.2%              ¦
¦Min/Maxpgahead    2      8    Accessed Virtual    0.3GB 27.6% Pinned     35.3%              ¦
+--------------------------------------------------------------------------------------------+
```

5.3.5 Performance Toolbox for AIX

The Performance Toolbox for AIX is a product for monitoring a group of systems. The Performance Toolbox for AIX can be used for the following actions:

Load monitoring Resource load must be monitored so that performance problems can be detected as they occur or (preferably) predicted well before they do.

Analysis and control When a performance problem is encountered, the proper tools must be selected and applied so that the nature of the problem can be understood and corrective action taken.

Capacity planning Long-term capacity requirements must be analyzed so that sufficient resources can be acquired well before they are required.

The Performance Toolbox for AIX consists of a Manager and Agents. The manager acts as a central repository for collecting performance data from each Agent.

Performance Toolbox Manager

The Manager component of the Performance Toolbox for AIX contains commands for collecting and analyzing performance data from the various agents. The following list shows the most common components:

xmperf The main user interface program providing a graphical display of local and remote performance information and a menu interface to commands of your choice.

3dmon A program that can monitor and display statistics in a 3-dimensional graph.

3dplay A program to play 3dmon recordings back in a 3dmon-like view.

chmon This program enables monitoring of vital statistics from a character terminal.

exmon This program enables monitoring of alarms generated by the filtd daemon running on remote hosts.

azizo or **jazizo** This program enables you to analyze long-term xmtrend recordings of performance data.

wlmperf A program for analyzing Workload Management activity from xmtrend recordings.

Performance Toolbox Agent

The Agent component of the Performance Toolbox for AIX contains a number of commands and applications for collecting data from various system resources. The following list shows the most common components:

xmservd The data-supplier daemon, which permits a system where this daemon runs to supply performance statistics to the Manager.

xmtrend Long-term recording daemon. Provides large metric set trend recordings for post-processing by jazizo.

filtd A daemon that can be used to do data reduction of existing statistics and to define alarm conditions and triggering of alarms.

xmpeek A program that enables you to display the status of xmservd on the local or a remote host and to list all available statistics from the daemon.

For detailed information about these commands, tools, and required fileset, refer to the appropriate manual pages or refer to *AIX 5L Practical Performance Tools and Tuning Guide*, SG24-6478.

5.3.6 LoadLeveler for AIX

LoadLeveler® for AIX schedules jobs and provides functions for building, submitting, and processing jobs quickly and efficiently in a dynamic environment.

Jobs are allocated to partitions in the cluster by a scheduler. The allocation of the jobs depends on the availability of resources within the cluster and various rules, which can be defined by the LoadLeveler administrator.

A user submits a job using a job command file, and the LoadLeveler scheduler attempts to find resources within the cluster to satisfy the requirements of the job. At the same time, it is the job of LoadLeveler to maximize the efficiency of the cluster. It attempts to do this by maximizing the utilization of resources, while at the same time minimizing the job turnaround time experienced by users.

5.4 Manage system software updates

In 3.5, “Installation of the operating system into a partition” on page 92, we introduce the operating system installation procedures. In this section, we discuss system documentation and ongoing software support. In addition, this section introduces a topic about the basic installation requirements for Linux.

5.4.1 Software packaging

Understanding this packaging will assist in the proper maintenance and support of the AIX 5L operating system.

Each software product can contain separately installable parts. The following list explains how AIX 5L Version 5.3 is organized:

Filesets The smallest installable base unit for the AIX 5L operating system. For example, bos.net.uucp.

Packages A group of separately installable filesets that provides a set of related functions. For example, bos.net.

Licensed Program Product
A complete software product including all packages associated with that licensed program. For example, BOS is a licensed program.

Bundle A list of software that can contain filesets, packages, and licensed program products (LPPs) put together for a particular use. For example, CDE, GNOME, and Netscape.

PTFs PTF is an acronym for program temporary fix. A PTF is an updated fileset or a fileset that fixes a previous system problem.

APAR An interim fix is a fix for a unique problem on the system. APARs will eventually become PTFs after testing and verification.

APARs are further divided into:

- Security APAR fix packages

  AIX 5L security advisory notifications per ITCS104. All requisites are included.

- Hiper APAR fix packages

  Individual "hiper" APARs for highly pervasive problems. All requisites are included.

- Critical APAR fix packages

  Individual packaging APARs for IBM-designated PTF collections that address combinations of hiper and security APARs. All requisites are included.

5.4.2 AIX 5L Version 5.3 installation pack

This section describes the physical installation package provided for installing AIX 5L Version 5.3. The packaging discussed in this document is based on AIX 5L Version 5.3 5765-G03 of 08/2004 and is subject to change with future releases.

You should receive the following CDs:

- *Volumes 1 to 8 of AIX 5L Version 5.3* CDs
  - The basic installation of AIX is on the *Volume 1* and *Volume 2* CDs. *Volume 1* contains mainly the boot image and the minimum programs required to have the operating system running. *Volume 1* also has basic device drivers for devices required to start the system.
  - Depending on the options selected with the installation, you might be prompted to insert *Volume 2*.
  - *AIX 5L V5.3 Volume 3* CD contains printer drivers and tools for system and performance management.
  - *AIX 5L V5.3 Volume 4* CD contains additional support for software.
  - *AIX 5L V5.3 Volumes 5-8* contain national language support (NLS) which is support for languages and regions outside the default installation language and regional support (us_english). *Volumes 5-6* include time zones and support for different currencies and regions.
- *Linux Toolbox* CD

  This delivery includes the RPM Package Manager (RPM), compilers (gcc and g++), and a substantial number of GNU utilities. The current list of available packages is available in the CONTENTS file in the top-level directory of the CD. If you chose to use GNOME and KDE desktop, you will be prompted to insert the *Linux Toolbox* CD.
- *Mozilla Web browser* CD

  This CD contains the Mozilla Web browser.
- *Documentation* CD

  This CD contains the Information Center.
- Expansion CD

  This CD includes useful *bundles* such as SSL or an HTTP server.

5.4.3 Service Update Management Assistant

Between installations and code releases, there are improvements, enhancements, and problems fixes. Problems that are found after a release are fixed with program temporary fixes (PTFs). These might be problems with the code or incompatibility for heterogeneous systems, such as running IBM products with products from other vendors. This section explains how to maintain your operating system after the basic operating system installation.

Classic method for software management

The old method of software management involved logging in to the Internet and finding fixes or upgrades. This method can still be used. You can access fixes for your operating systems using the Fix Central Web site:

http://www.ibm.com/eserver/support/fixes/

From the Web site, follow these steps:

1. From the Server list, select the appropriate family, for example, the iSeries family or pSeries family.
2. From the Product or fix type list, select the operating system for which you want to get a fix.
3. Depending on your selections for Server and Product or fix type, you might see additional lists from which you can select specific options.
4. Click **Continue**.

For more information about fixes for the operating systems, see the following Web sites:

- AIX 5L and Linux

  http://www.ibm.com/servers/eserver/support/unixservers/index.html

- i5/OS

  http://publib.boulder.ibm.com/infocenter/iseries/v5r3/ic2924/index.htm?info/rzam8/rzam81.htm

New preferred method for software management

Using the previous method can be a time-consuming exercise for the system administrator. Through the **compare_report** command and its SMIT interface, AIX 5L Version 5.2 and later enables you to compare installed software or fix repositories to a list of available fixes from the IBM Support Web site. This enables system administrators to develop a proactive fix strategy.

AIX 5L Version 5.3 introduces automatic download, scheduling, and notification capabilities through the new *Service Update Management Assistant (SUMA)* tool. SUMA is fully integrated into the AIX 5L Base Operating System. SUMA provides unattended task-based download of APARs, PTFs, and recommended maintenance levels (MLs). SUMA can also be configured to periodically check the availability of specific new fixes and entire maintenance levels. All SUMA modules and the **suma** command are contained in the bos.suma fileset. The **lslpp -p bos.suma** command can be used to verify the requisites for the bos.suma fileset:

```
# lslpp -p bos.suma
  Fileset               Requisites
  ----------------------------------------------------------------------------
Path: /usr/lib/objrepos
  bos.suma 5.3.0.0      *prereq bos.rte 5.1.0.0
```

Functional description

SUMA is implemented as a task-oriented utility that supports the following features:

- Automated task-based retrieval of the following fix types and categories:
  - Specific APAR
  - Specific PTF
  - Latest critical PTFs
  - Latest security PTFs
  - All latest PTFs
  - Specific fileset
  - Specific maintenance level
- Three different task actions to initiate download preview, code download, or combined download and fix repository cleanup (using the **lppmgr** command).
- Support for FTP, HTTP, or HTTPS transfer protocols and proxy servers. (HTTPS requires OpenSSL to be installed from the AIX Toolbox for Linux Applications.
- E-mail notification of update availability and task completion.
- Messaging function providing six verbosity levels (Off, Error, Warning, Information, Verbose, and Debug) for sending information to the screen, log file, or e-mail address.

Concepts and implementation specifics

The SUMA controller uses certain SUMA modules to execute SUMA operations and functions. Figure 5-2 on page 154 shows module dependencies.

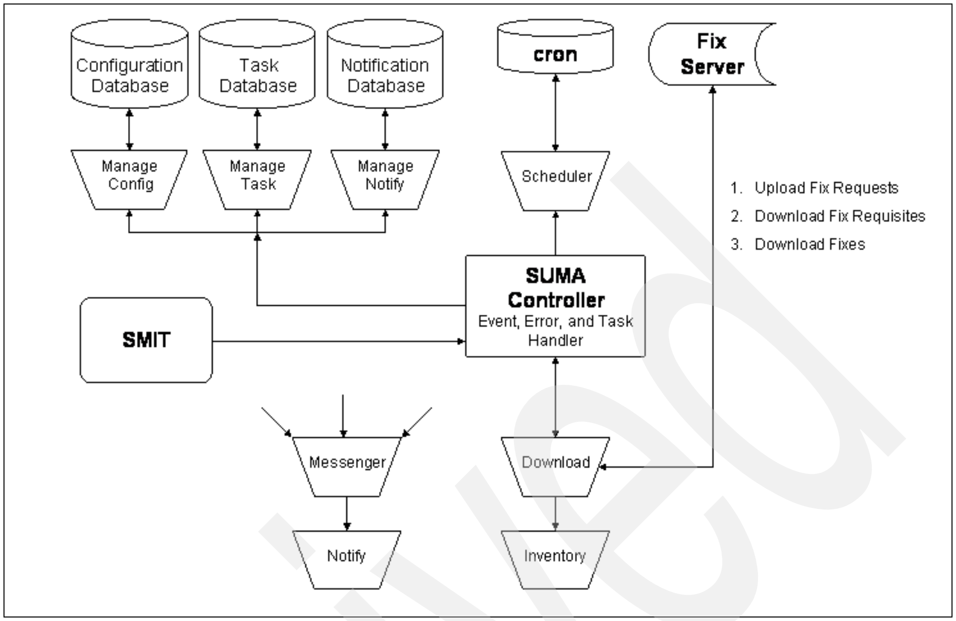


Figure 5-2 SUMA module dependencies

The SUMA modules supply the following services and functions:

- Download module

  Provides functions related to network activities and is solely responsible for communicating with the IBM @server pSeries support server.

- Manage configuration module

  Represents a utility class containing global configuration data and general-purpose methods.

- Messenger module

  Provides messaging, logging, and notification capabilities. There are two log files located in the /var/adm/ras/ directory. The /var/adm/ras/suma.log log file contains any messages that pertain to SUMA controller operations. The other log file, /var/adm/ras/suma_dl.log, tracks the download history of SUMA download operations and contains entries of the form DateStamp:FileName.

- Notify module

  Manages the file that holds the contact information for event notifications.

- Task module

  Creates, retrieves, views, modifies, and deletes SUMA tasks.

- Scheduler module

  Handles scheduling of SUMA task execution and interacts with the AIX **cron** daemon.

- Inventory module

  Returns the software inventory (installed or in a repository) of the local system (localhost) or a NIM client.

- Utility and database modules

  These are additional modules that supply private utilities for SUMA code and utilities for handling the stanza-style SUMA databases.

Command line interface

The **suma** command provides task-related SUMA control functions. The command can be used to perform the following operations on a SUMA task:

- Create
- Edit
- List
- Schedule
- Unschedule
- Delete

The usage information of the **suma** command is as follows:

```
# suma -?
Usage:
Create, Edit, or Schedule a SUMA task.
        suma { { [-x][-w] } | -s CronSched } [ -a Field=Value ]... [ TaskID ]
```

SMIT user interface

SUMA related tasks and functions are supported by SMIT menus and panels. The new main **suma** menu shown in Example 5-9 on page 156 can be directly accessed through the SMIT fast path **smitty suma**.

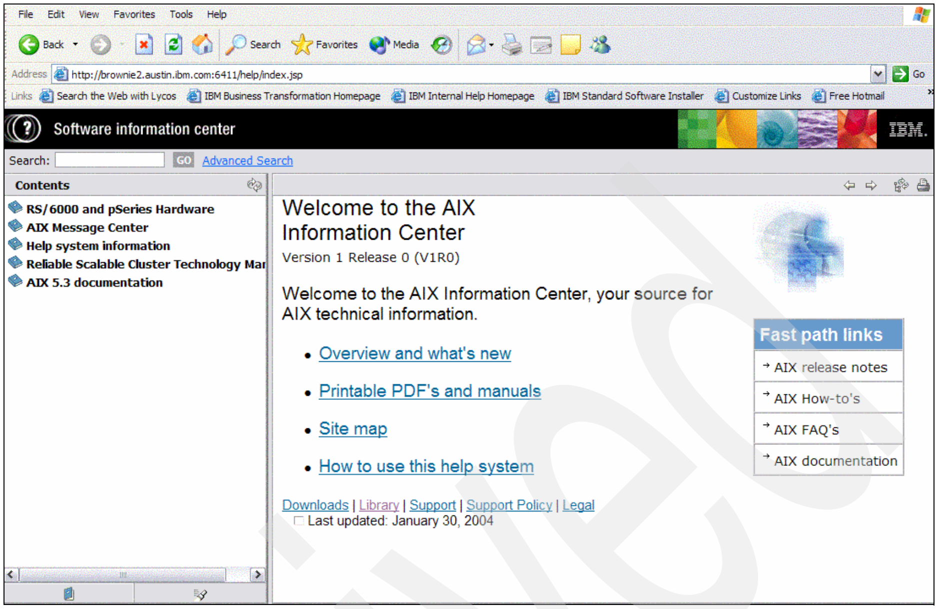
Example 5-9 Smitty main menu for SUMA operations

```
                   Service Update Management Assistant (SUMA)

Move cursor to desired item and press Enter.

  Download Updates Now (Easy)
  Custom/Automated Downloads (Advanced)
  Configure SUMA


Esc+1=Help          Esc+2=Refresh       Esc+3=Cancel        Esc+8=Image
Esc+9=Shell         Esc+0=Exit          Enter=Do
```

Using the download options of this menu and closely examining the SMIT command output will give some insight into the way SUMA retrieves updates from the IBM eserver Support Web site.

5.4.4 System documentation

The *AIX 5L Version 5.3 Documentation* CD contains AIX 5L documentation including online help facilities and the information center. Starting with AIX 5L Version 5.3, IBM eserver pSeries and AIX 5L documentation is available in one of two information centers: the IBM eserver pSeries and AIX Information Center on the Web and the AIX Information Center on the documentation CD. Figure 5-3 on page 157 shows the information center.


File Edit View Favorites Tools Help
Back
Search
Favorites
Media
Address http://brownie2.austin.ibm.com:6411/help/index.jsp
Go
Links Search the Web with Lycos IBM Business Transformation Homepage IBM Internal Help Homepage IBM Standard Software Installer Customize Links Free Hotmail
Software information center
IBM.
Search:
GO
Advanced Search
Contents
RS/6000 and pSeries Hardware
AIX Message Center
Help system information
Reliable Scalable Cluster Technology Mar
AIX 5.3 documentation
Fast path links
AIX release notes
AIX How-to's
AIX FAQ's
AIX documentation
Overview and what's new
Printable PDF's and manuals
Site map
How to use this help system
Downloads | Library | Support | Support Policy | Legal

Welcome to the AIX Information Center

Version 1 Release 0 (V1R0)

Welcome to the AIX Information Center, your source for AIX technical information.

Last updated: January 30, 2004

Figure 5-3 AIX Information Center on the documentation CD

The AIX and pSeries Information Center is more than a portal to documentation. From this Web site, you can access the following tools and resources:

http://publib16.boulder.ibm.com/pseries/en_US/infocenter/base/

- A message database that shows what error messages mean and, in many cases, how you can recover. The database also includes LED codes.
- How-to tips with step-by-step instructions for completing system administrator and user tasks.
- FAQs for quick answers to common questions.
- The entire AIX 5L software documentation library for Version 5.1, Version 5.2, and Version 5.3. Each publication is available in PDF format, and abstracts are provided for books for Version 5.2 and Version 5.3.
- Centralized information previously located throughout the library and easier access to information about some new AIX 5L functions:
  - A new selection in the navigation bar centralizes all partitioning information, including planning, installation, and implementation information for partitioned system operations.
  - A new *Advanced Accounting* publication is available. *Advanced Accounting* offers increased flexibility, enabling users to customize it to meet their needs.

- A new *Partition Load Manager for AIX Guide and Reference* provides experienced system administrators with information about how to perform such tasks as installing, configuring, and managing the Partition Load Manager for AIX.
- Links to the entire pSeries and p5 hardware documentation library.
- A resources page that links users to other IBM and non-IBM Web sites proven useful to system administrators, application developers, and users.
- Links to related documentation from IBM, including white papers, IBM Redbooks, and technical reports.
- Several new videos are available for client-installable features and client-replaceable parts.

For complete information about setting up and starting up the information center, see the following document:

http://publib.boulder.ibm.com/infocenter/pseries/topic/com.ibm.aix.doc/aixins/insgdrf/insgdrf.pdf

5.5 Linux introduction

Linux is one of the operating systems that can be installed on your server or logical partition. This section provides information about how to prepare for installing Linux on your system, how to install a Linux for POWER distribution, and how to install software to enable dynamic logical partitioning and hot plug capabilities.

5.5.1 Planning for Linux

Before installing Linux on an IBM @server hardware system, determine whether you want to install Linux with or without logical partitions.

It is essential that you understand Linux concepts before you start creating Linux logical partitions. The purpose of this information is to familiarize you with the hardware and software required for Linux logical partitions. For the most current information about supported hardware devices for Linux, refer to the following link:

http://www.ibm.com/servers/eserver/pseries/hardware/factsfeatures.html

5.5.2 Minimum configuration requirements for a Linux partition

In order to use features that require logical partitions on an IBM @server OpenPower system, you need the appropriate Advanced OpenPower Virtualization technologies activated in your system's Hardware Management Console.

Each Linux logical partition on an IBM @server hardware system requires the following minimum hardware resources:

- 1 dedicated processor or 0.1 processing unit
- 128 MB
- Storage adapter (physical or virtual)
- Network adapter (physical or virtual)
- Approximately 1 GB storage

5.5.3 Shared processor support for Linux logical partitions

The POWER5 processor-based systems provide shared processor support that allows Linux to run on less than 100% of a processor. The minimum number of shared processing units that can be allocated to a Linux partition is 0.1.

On IBM @server i5, p5, and OpenPower systems, the system administrator can alter the shared processor units for a Linux for POWER distribution with the Linux 2.6 kernel and the Dynamic Resource Manager (DRM) without restarting the operating system by using a dynamic LPAR operation. On a Linux for POWER distribution with the Linux 2.4 kernel, changing the shared processing units allocated to the partition requires the operating system to be restarted. On Linux on iSeries systems, the system administrator can alter the shared processor units without involving Linux.

5.5.4 Installing a Linux distribution

You can find installation information for both Red Hat Enterprise Linux Version 3 and for SUSE LINUX Enterprise Server 9 in the documentation provided on the Linux installation CDs.

For detailed information about installing Red Hat Enterprise Linux Version 3 on POWER hardware, see *Red Hat Enterprise Linux 3 Installation Guide for the IBM @server iSeries and IBM @server pSeries Architectures* at:

http://www.redhat.com/docs/manuals/enterprise/RHEL-3-Manual/ppc-multi-install-guide/

For detailed information about installing SUSE LINUX Enterprise Server 9, see the installation information at:

http://www.novell.com/documentation/sles9/index.html

We highly recommended that you use the additional software for Linux topic to enhance your Linux on POWER solutions. Visit the following link for more information:

http://techsupport.services.ibm.com/server/lopdiags

5.5.5 Virtual I/O Server and Linux

The Virtual I/O Server provides virtual storage and shared Ethernet capability to client logical partitions on the system. It allows physical adapters with attached disks on the Virtual I/O Server logical partition to be shared by one or more client partitions. Virtual I/O Server partitions are not intended to run applications or for general user logins. The Virtual I/O Server is installed in its own logical partition.

The Virtual I/O Server supports client logical partitions running the following Linux operating systems:

- SUSE LINUX Enterprise Server 9 for POWER
- Red Hat Enterprise Linux AS for POWER Version 3

5.5.6 Dynamically managing physical I/O devices and slots on Linux

A Linux distribution with Linux kernel Version 2.6 or later is required in order to dynamically move I/O devices and slots to or from a Linux logical partition.

If you add slots with adapters, the devices are automatically configured by Linux kernel modules (rpaphp and PCI Hotplug Core). However, after the devices have been added with the HMC, you must log in to the running Linux logical partition as root so that you can set up those devices that have been added using the appropriate user space tools, such as the **mount** command or the **ifup** command.

If you remove adapters for storage devices, you must unmount the file systems on those devices before you remove the slots and adapters. Also, if you remove network adapters, you should shut down the network interfaces for those devices before removing the slots and adapters.

5.6 Basic problem determination

This section describes the error handling of the eserver p5 system and how to capture service data for IBM service representatives.

5.6.1 Error analyzing

Because the service processor monitors the hardware environmental and FFDC (FIR bits) activities, it is the primary collector of platform hardware errors and is used to begin the analysis and processing of these events. The service processor will identify and sort errors by type and criticality. In effect, the service processor initiates a preliminary error analysis to categorize events into specific categories:

- Errors that are recoverable but should be recorded for threshold monitoring. These events do not require immediate service but should be logged and tracked to look for, and effectively respond to, future problems.
- Fatal system errors (initiate server reboot and IPL, error analysis, and call home if enabled).
- Recoverable errors that require service either because an error threshold has been reached or a component has been taken offline (even if a redundant component has been used for sparing).

When a recoverable and serviceable error (the third type) is encountered, the service processor notifies the POWER Hypervisor, which places an entry into the operating system error log. The operating system log contains all recoverable error logs. These logs represent either recoverable platform errors or errors detected and logged by I/O device drivers. Operating system error log analysis (ELA) routines monitor this log, identify serviceable events (ignoring information-only log entries), and copy them to a diagnostic event log. At this point, the operating system sends an error notification to a client-designated user (by default, the root user). This action also invokes the Electronic Service Agent application, which initiates appropriate system serviceability actions:

- On servers that do not include an HMC, the Electronic Service Agent notifies the system operator of the error condition and, if enabled, also initiates a call for service. The service call can be directed to the IBM support organization, or to a client-identified pager or server identified and set up to receive service information. Note that an HMC is required on p5-590 and p5-595 servers.
- On servers equipped with an HMC (including the p5-595 and p5-590 servers), the Electronic Service Agent forwards the results of the diagnostic error log analysis to the Service Focal Point application running on the HMC. The Service Focal Point consolidates and reports errors to IBM or a user-designated system or pager.
- The AIX conslog provides a log of boot messages. All messages can be found in /var/adm/ras/conslog.

In either case, the failure information includes:

- The source of the error

- The part numbers of the components needing repair
- The location of those components
- Any available extended error data

The failure information is sent back to IBM service for parts ordering and additional diagnosis if required. This detailed error information enables IBM service representatives to bring probable replacement hardware components when a service call is placed, minimizing system repair time.

In a multisystem configuration, any HMC-attached @server p5 server can be configured to forward call home requests to a central Electronic Service Agent Gateway (SAG) application on an HMC that owns a modem and performs the call home on behalf of any of the servers.

Figure 5-4 shows the error reporting structure of a POWER5 environment.

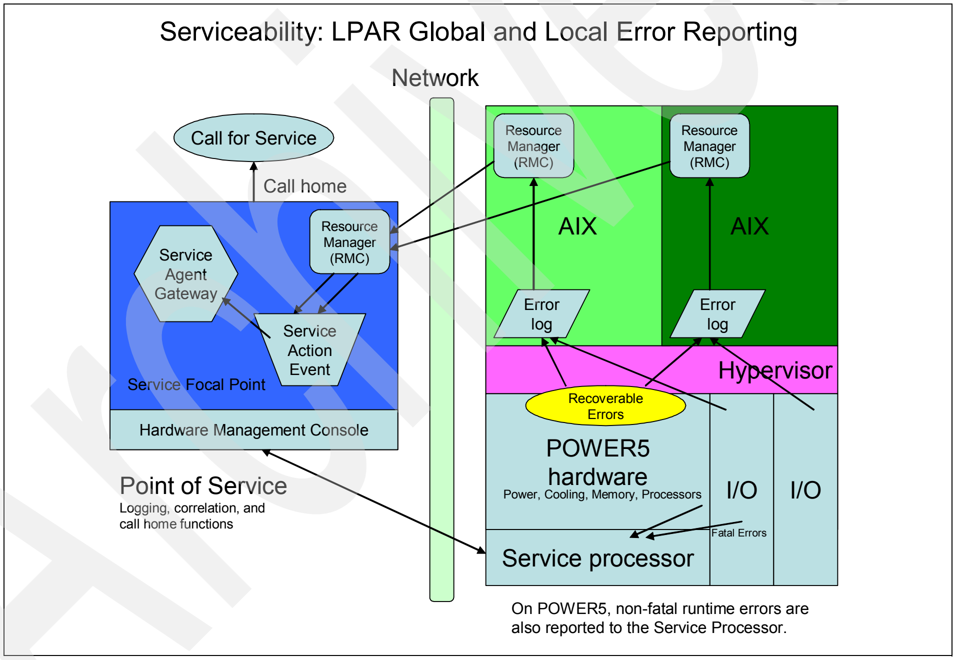


Figure 5-4 Error reporting structure of a POWER5 environment

5.6.2 Managing the error log

The following sections discuss various error log management-related functions. In addition to the AIX conslog, the error log is an important repository for all system-related errors.

Configuring an error log file (/var/adm/ras/errlog)

This section discusses the use of the **errdemon** command to customize the error log file:

- To list the current values for the error log file name, error log file size, and buffer size that are currently stored in the error log configuration database settings, use the following command:

  ```
  # /usr/lib/errdemon -l
  Error Log Attributes
  --------------------------------------------
  Log File                /var/adm/ras/errlog
  Log Size                4096 bytes
  Memory Buffer Size      8192 bytes
  ```

  The preceding example shows the default error log file attributes. The log file size cannot be made smaller than the hard-coded default of 4 KB, and the buffer cannot be made smaller than the hard-coded default of 8 KB.

- To change the log file name to /var/adm/ras/errlog.test, use the following command:

  ```
  /usr/lib/errdemon -i /var/adm/ras/errlog.test
  ```

- To change the log file size to 8 KB, use the following command:

  ```
  /usr/lib/errdemon -s 8192
  ```

  If the log file size specified is smaller than the size of the log file currently in use, the current log file is renamed by appending .old to the file name and a new log file is created with the specified size limit. The amount of space specified is reserved for the error log file and is not available for use by other files. Therefore, be careful not to make the log excessively large. But, if you make the log too small, important information might be overwritten prematurely. When the log file size limit is reached, the file wraps. That is, the oldest entries are overwritten by new entries.

- To change the size of the error log device driver's internal buffer to 16 KB, use the following command:

  ```
  # /usr/lib/errdemon -B 16384
  0315-175 The error log memory buffer size you supplied will be rounded up
  to a multiple of 4096 bytes.
  ```

  If the specified buffer size is larger than the buffer size currently in use, the in-memory buffer is immediately increased, and if the specified buffer size is smaller than the buffer size currently in use, the new size is put into effect the next time the error daemon is started after the system is rebooted. The size you specify is rounded up to the next integral multiple of the memory page size (4 KB).

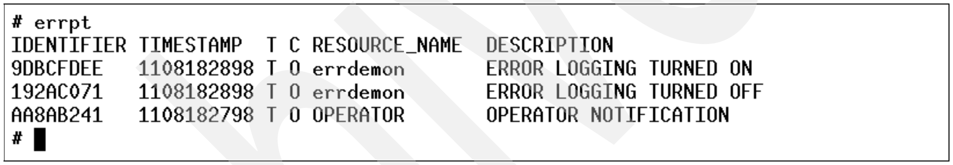
You should be careful not to impact your system's performance by making the buffer excessively large. But, if you make the buffer too small, the buffer might become full if error entries are arriving faster than they are being read from the buffer and put into the log file. When the buffer is full, new entries are discarded until space becomes available in the buffer. When this situation occurs, an error log entry is created to inform you of the problem.

The following command shows the new attributes of the error log file:

```
# /usr/lib/errdemon -l
Error Log Attributes
--------------------------------------------
Log File                /var/adm/ras/errlog.test
Log Size                8192 bytes
Memory Buffer Size      16384 bytes
```

Starting the error logging daemon

To determine if error logging daemon is on or off, issue the **errpt** command. The **errpt** command output can contain entries, as shown in Figure 5-5.



```
# errpt
IDENTIFIER TIMESTAMP  T C RESOURCE_NAME  DESCRIPTION
9DBCFDEE   1108182898 T O errdemon       ERROR LOGGING TURNED ON
192AC071   1108182898 T O errdemon       ERROR LOGGING TURNED OFF
AA8AB241   1108182798 T O OPERATOR       OPERATOR NOTIFICATION
# 
```

Figure 5-5 Sample errpt command output

If the **errpt** command does not generate entries, error logging has been turned off. To activate the daemon, use the following command:

```
/usr/lib/errdemon
```

The errdemon daemon starts error logging and writes error log entries in the system error log.

Stopping the error logging daemon

To stop the error logging daemon from logging entries, use the following command:

```
/usr/lib/errstop
```

Cleaning an error log

Cleaning the error log implies deleting old or unnecessary entries from the error log. Cleaning is normally done as part of the daily **cron** command execution. If it is not done automatically, you should probably clean the error log regularly.

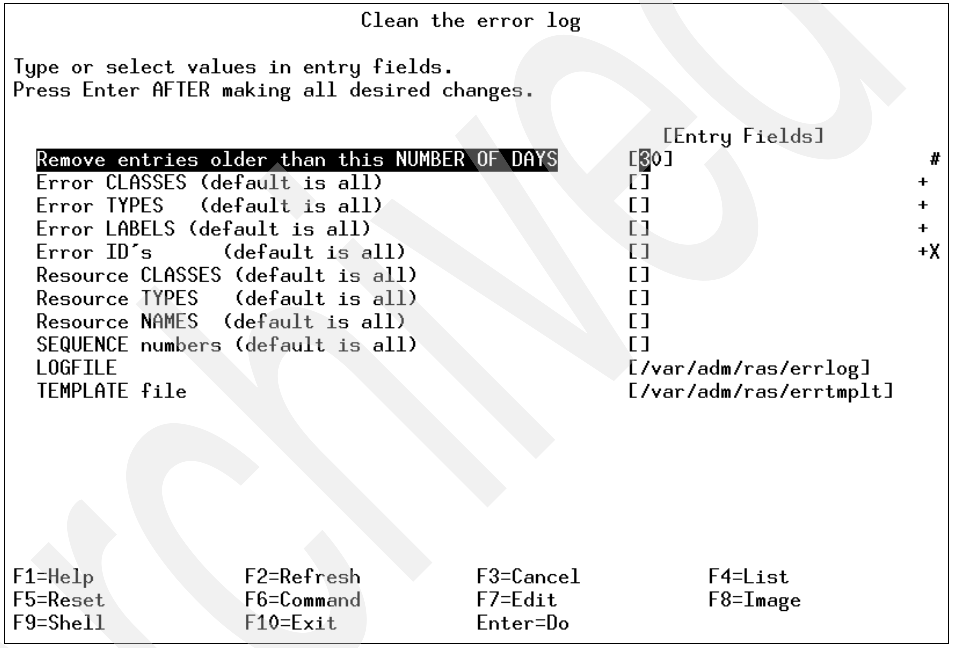
To delete all the entries from the error log, use the following command:

```
errclear 0
```

To selectively remove entries from the error log, for example, to delete all software errors entries, use the following command:

```
errclear -d S 0
```

Alternatively, use the SMIT fast path command (**smit errclear**), which will display the screen shown in Figure 5-6. Fill in the appropriate fields, as per your requirements, to clean the error log.



```
                              Clean the error log

Type or select values in entry fields.
Press Enter AFTER making all desired changes.

                                                         [Entry Fields]
  Remove entries older than this NUMBER OF DAYS      [30]                    #
  Error CLASSES (default is all)                     []                     +
  Error TYPES   (default is all)                     []                     +
  Error LABELS (default is all)                      []                     +
  Error ID's      (default is all)                   []                     +X
  Resource CLASSES (default is all)                  []
  Resource TYPES   (default is all)                  []
  Resource NAMES  (default is all)                   []
  SEQUENCE numbers (default is all)                  []
  LOGFILE                                            [/var/adm/ras/errlog]
  TEMPLATE file                                      [/var/adm/ras/errtmplt]




F1=Help             F2=Refresh          F3=Cancel           F4=List
F5=Reset            F6=Command          F7=Edit             F8=Image
F9=Shell            F10=Exit            Enter=Do
```

Figure 5-6 Clean the error log menu

Generating an error report

The **errpt** command generates the default summary error report that contains one line of data for each error. It includes flags for selecting errors that match specific criteria. By using the default condition, you can display error log entries in the reverse order they occurred and were recorded. By using the **-c** (concurrent) flag, you can display errors as they occur. You can use flags to generate reports with different formats. Example 5-10 on page 166 shows the syntax of the **errpt** command.

Example 5-10 Syntax of the errpt command

```
errpt [ -a ] [ -c ] [ -d ErrorClassList ] [ -e EndDate ] [ -g ] [ -i File ]
[ -j ErrorID [ ,ErrorID ] ] | [ -k ErrorID [ ,ErrorID ] ] [ -J ErrorLabel
[ ,ErrorLabel ] ] | [ -K ErrorLabel [ ,ErrorLabel ] ] [ -l SequenceNumber ]
[ -m Machine ] [ -n Node ] [ -s StartDate ] [ -F FlagList ]
[ -N ResourceNameList ] [ -R ResourceTypeList ] [ -S ResourceClassList ]
[ -T ErrorTypeList ] [ -y File ] [ -z File ]
```

Table 5-1 provides the flags commonly used with the **errpt** command.

Table 5-1 Common flags for the errpt command

Flag	Description
-a	Displays information about errors in the error log file in detailed format.
-e *EndDate*	Specifies all records posted prior to and including the EndDate variable, where the EndDate variable has the form *mmddhhmmyy* (month, day, hour, minute, and year).
-J *ErrorLabel*	Includes the error log entries specified by the ErrorLabel variable. The ErrorLabel variable values can be separated by commas or enclosed in double-quotation marks and separated by commas or space characters.
-j *ErrorID [,ErrorID]*	Includes only the error log entries specified by the ErrorID (error identifier) variable. The ErrorID variables can be separated by commas or enclosed in double-quotation marks and separated by commas or space characters.
-K *ErrorLabel*	Excludes the error log entries specified by the ErrorLabel variable.
-k *ErrorID*	Excludes the error log entries specified by the ErrorID variable.
-N *ResourceNameList*	Generates a report of resource names specified by the ResourceNameList variable, where ResourceNameList is a list of names of resources that have detected errors.
-s *StartDate*	Specifies all records posted on and after the StartDate variable, where the StartDate variable has the form *mmddhhmmyy* (month, day, hour, minute, and year).

The output of **errpt** command without any flags displays the error log entries with the following fields:

IDENTIFIER Numerical identifier for the event.

TIMESTAMP Date and time of the event occurrence.

T Type of error. Depending on the severity of the error, the the possible error types are as follows:

- **PEND** The loss of availability of device or component is imminent.
- **PERF** The performance of the device or component has degraded to below an acceptable level.
- **PERM** Most severe errors, due to a condition that could not be recovered.
- **TEMP** Condition that was recovered after a number of unsuccessful attempts.
- **UNKN** Not possible to determine the severity of an error.
- **INFO** This is an informational entry.

C Class of error. The possible error classes are as follows:

- **H** Hardware.
- **S** Software.
- **O** Informational message.
- **U** Undetermined.

RESOURCE_NAME Name of the failing resource.

DESCRIPTION Summary of the error.

Using the **errpt** command without a flag outputs all the entries in the log, as shown in Example 5-11. Because the number of error log entries can exceed a single page, you can use **errpt | pg** to have a page-wise view.

Example 5-11 Output of the errpt command without a flag

```
# errpt
IDENTIFIER TIMESTAMP  T C RESOURCE_NAME  DESCRIPTION
0BA49C99   1106154298 T H scsi0          SCSI BUS ERROR
E18E984F   1106110298 P S SRC            SOFTWARE PROGRAM ERROR
2A9F5252   1106094298 P H tok0           WIRE FAULT
E18E984F   1102120498 P S SRC            SOFTWARE PROGRAM ERROR
E18E984F   1101105898 P S SRC            SOFTWARE PROGRAM ERROR
AD331440   1101104498 U S SYSDUMP        SYSTEM DUMP
E18E984F   1030182798 P S SRC            SOFTWARE PROGRAM ERROR
E18E984F   1030182698 P S SRC            SOFTWARE PROGRAM ERROR
E18E984F   1030182598 P S SRC            SOFTWARE PROGRAM ERROR
E18E984F   1023175198 P S SRC            SOFTWARE PROGRAM ERROR
E18E984F   1023175098 P S SRC            SOFTWARE PROGRAM ERROR
E18E984F   1023174898 P S SRC            SOFTWARE PROGRAM ERROR
```

```
2A9F5252   1022143498 P H tok0          WIRE FAULT
35BFC499   1022081198 P H hdisk0        DISK OPERATION ERROR
AD331440   1021185998 U S SYSDUMP       SYSTEM DUMP
0BA49C99   1021185798 T H scsi0         SCSI BUS ERROR
35BFC499   1021180298 P H hdisk0        DISK OPERATION ERROR
```

The preceding example shows that the resource name for a disk operation error is hdisk0. To obtain all the errors with the resource name hdisk0 from the error log, use the **errpt** command with the **-N** flag, as shown in Example 5-12.

Example 5-12 The errpt command with the -N flag

```
# errpt -N hdisk0
IDENTIFIER TIMESTAMP  T C RESOURCE_NAME  DESCRIPTION
35BFC499   1022081198 P H hdisk0        DISK OPERATION ERROR
35BFC499   1021180298 P H hdisk0        DISK OPERATION ERROR
```

Reading an error log report

To read the contents of an error log report, use the **errpt** command with the **-a** flag. For example, to read an error log report with the resource name hdisk0, use the **errpt** command with the **-a** and **-j** flags with an error identifier, as shown in the sample error report in Example 5-13.

Example 5-13 Sample error report

```
# errpt -a -j 35BFC499
--------------------------------------------------------------------------
LABEL:          DISK_ERR3
IDENTIFIER:     35BFC499

Date/Time:       Thu Oct 22 08:11:12
Sequence Number: 36
Machine Id:      006151474C00
Node Id:         sv1051c
Class:           H
Type:            PERM
Resource Name:   hdisk0
Resource Class:  disk
Resource Type:   scsd
Location:        04-B0-00-6,0
VPD:
        Manufacturer................IBM
        Machine Type and Model......DORS-32160    !#
        FRU Number..................
        ROS Level and ID............57413345
        Serial Number...............5U5W6388
        EC Level....................85G3685
        Part Number.................07H1132
        Device Specific.(Z0)........000002028F00001A
        Device Specific.(Z1)........39H2916
        Device Specific.(Z2)........0933
        Device Specific.(Z3)........1296
        Device Specific.(Z4)........0001
        Device Specific.(Z5)........16

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
SENSE DATA
0A06 0000 2800 0000 0088 0002 0000 0000 0200 0200 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 0001 0001 2FC0
```

In this example, the **errpt** output indicates that the error was hardware in nature and a PERM type, or severe error from which the system could not recover. This error was a disk operation error on hdisk0 and might be due to a failing disk drive or a bad device cable. Given that the error appeared twice in the **errpt** output, it might be necessary to have a customer engineer service the machine by checking cables and hardware, replacing any parts that have worn out.

Copying an error log to diskette or tape

You might need to send the error log to the AIX System Support Center for analysis. To copy the error log to a diskette, place a formatted diskette into the diskette drive and use the following commands:

```
ls /var/adm/ras/errlog | backup -ivp
```

To copy the error log to tape, place a tape in the drive and enter:

```
ls /var/adm/ras/errlog | backup -ivpf/dev/rmt0
```

Log maintenance activities

The **errlogger** command enables a system administrator to record messages in the error log. Whenever you perform a maintenance activity, replace hardware, or apply a software fix, it is a good idea to record this activity in the system error log.

The following example shows the log entries before and after a message (Error Log cleaned) was logged by an operator (using the **errlogger** command) in the error log.

Example 5-14 Log entries before and after a message

```
# errpt
IDENTIFIER TIMESTAMP  T C RESOURCE_NAME  DESCRIPTION
9DBCFDEE   1109164598 T O errdemon       ERROR LOGGING TURNED ON
192AC071   1109164598 T O errdemon       ERROR LOGGING TURNED OFF
# errlogger "Error Log cleaned"
# errpt
IDENTIFIER TIMESTAMP  T C RESOURCE_NAME  DESCRIPTION
AA8AB241   1109164698 T O OPERATOR       OPERATOR NOTIFICATION
9DBCFDEE   1109164598 T O errdemon       ERROR LOGGING TURNED ON
192AC071   1109164598 T O errdemon       ERROR LOGGING TURNED OFF
```

The identifier AA8AB241 in the preceding example is the message record entry with a description of OPEARATOR NOTIFICATION.

5.6.3 Capturing service data

A useful command to gather system data for an IBM representative is the **snap** command.

The **snap** command gathers system configuration information and compresses the information into a PAX file. The file can then be written to a device, such as tape or DVD, or transmitted to a remote system. The information gathered with the **snap** command might be required to identify and resolve system problems.

> **Note:** Root user authority is required to execute the **snap** command. Use the **snap -o /dev/cd0** command to copy the compressed image to DVD. Use the **snap -o /dev/rmt0** command to copy the image to tape.

The **snap -g** command gathers general system information, including the following information:

- Error report
- Copy of the customized Object Data Manager (ODM) database
- Trace file

- User environment
- Amount of physical memory and paging space
- Device and attribute information
- Security user information

The output of the **snap -g** command is written to the /tmp/ibmsupt/general/general.snap file.

The **snap** command checks for available space in the /tmp/ibmsupt directory, the default directory for **snap** command output.

For detailed information about the **snap**, refer to the appropriate manual pages.

5.7 Logical volume management

This section describes the basics of disk management. This is the basic knowledge of the Logical Volume Manager (LVM). For more information, consult *AIX Logical Volume Manager from A to Z: Introduction and Concepts*, SG24-5432, available at:

http://www.redbooks.ibm.com/abstracts/sg247039.html

5.7.1 LVM introduction

This topic provides an overview of the LVM concepts. Figure 5-7 on page 172 contains a basic description of logical volume management.

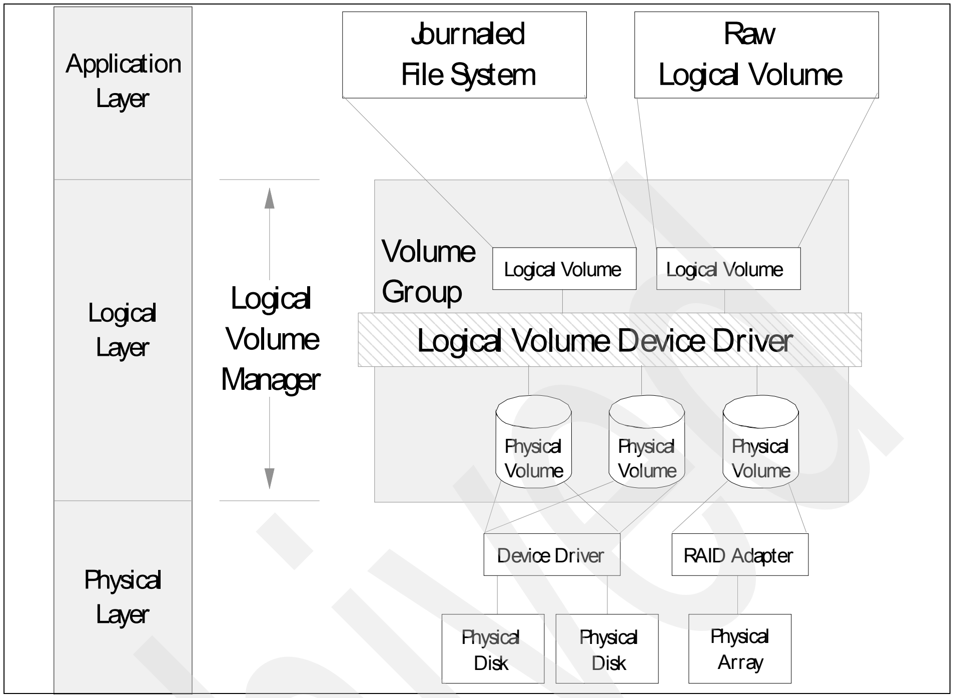


Figure 5-7 How LVM entities fit together

At the physical layer of the LVM are physical disks (PVs). The physical disks consist of small partitions called physical partitions (PPs). A physical partition is a smallest allocatable unit of the physical disk. A PV is not usable unless it belongs to a volume group (VG). A volume group contain one or more physical volumes. The total size of the volume group is the sum of all physical partitions in all the physical volumes in the volume group. To access a physical volume, logical volumes (LVs) must be created. Logical volumes consist of logical partitions (LPs) that directly map to physical partitions on the physical disks. Logical volumes are used by applications either as raw devices or as file systems. The file system and the raw devices are used by applications to access the disks. The software that manages the creation and usage is called the Logical Volume Manager (LVM), which employs the Logical Volume Device Driver (LVDD) to submit the read/write requests to the physical disks.

5.7.2 Physical volume management

This section describes physical volume management.

Before adding a disk, answer the following questions:

- Is the physical volume a real hard disk or an exported logical volume from a Virtual I/O Server?
- Can the physical volume be replaced online (hot plug)?
- Is it used as a hot spare?

Hot-spare disks

Beginning with AIX 5L Version 5.1, you can designate disks as hot-spare disks for a volume group with mirrored logical volumes. You must also specify a policy to be used if a disk or disks start to fail. You must also specify synchronization characteristics. A hot-spare disk must at least have the same capacity as the smallest disk in the volume group. The use of a hot spare is to wait for a disk in the volume group to have failures above the acceptable threshold based on the specified policies and replace the disk when the threshold is reached.

The **chvg** and **chpv** commands enable hot-spare disk support. They provide several options used for the hot-spare features. The **chpv -h** command sets a physical volume as a *hot spare* and the **chvg** command specifies the hot-spare policies and synchronization policy. It has the following syntax:

```
chvg -hhotsparepolicy -ssyncpolicy VolumeGroup
```

The hot-spare policy determines how and when the hot spare should take over the failing disk. The synchronization policy determines when and how to synchronize the new disk.

There are some changes in the output of the **lsvg** and **lspv** commands to reflect the hot-spare features. Example 5-15 shows the **lsvg** command output changes.

Example 5-15 The lsvg command output with hotspare at the bottom

```
# lsvg rootvg
VOLUME GROUP:       rootvg                   VG IDENTIFIER:  00cc489e00004c00000
0010256185b3f
VG STATE:           active                   PP SIZE:        64 megabyte(s)
VG PERMISSION:      read/write               TOTAL PPs:      542 (34688 megabytes)
MAX LVs:            256                      FREE PPs:       403 (25792 megabytes)
LVs:                12                       USED PPs:       139 (8896 megabytes)
OPEN LVs:           9                        QUORUM:         2
TOTAL PVs:          1                        VG DESCRIPTORS: 2
STALE PVs:          0                        STALE PPs:      0
ACTIVE PVs:         1                        AUTO ON:        yes
MAX PPs per VG:     32512                                    0
MAX PPs per PV:     1016                     MAX PVs:        32
LTG size (Dynamic): 256 kilobyte(s)          AUTO SYNC:      no
HOT SPARE:          no                       BB POLICY:      relocatable
```

5.7.3 Adding a physical disk

This section helps with the basic steps to add disks to your system. Use one of the following methods to add disks to the system.

Method 1

This method is suitable for non-hot plug devices and it is possible to get down time from the client. Perform the following steps:

1. Shut down the system.
2. Add the disk.
3. Restart the system. If the disk is added to the system, it will be configured by **`cfgmgr`** when the system runs its **`rc.boot`** 3 step.
4. Add the disk to a volume group using the **`extendvg`** command.

Method 2

This method is suitable for hot plug disks. Perform the following steps:

1. At the command line, type **`diag`** and press Enter.
2. On the Function Selection window, select **Task Selection**.
3. On the Tasks Selection window, select **Hot Plug Task**.
4. Select **SCSI** and **SCSI RAID Hot Plug Manager**.
5. Select **Attach a Device to a SCSI Hot Swap Enclosure Device**.
6. A list of empty slots in the SCSI hot swap enclosure device is shown.
7. Select the slot where you want to install the disk drive and press Enter.
8. Run the **`cfgmgr`** (or **`mkdev -l`**) command to configure the device, and use the **`lspv`** command to see the newly configured disk.
9. Add the disk to a volume group using the **`extendvg`** command.

> **Note:** This procedure uses the **`diag`** command to insert the disk. The **`diag`** command is the recommended method. It is possible to use the command line for this procedure. The **`lsslot`** command identifies hot-plug slots while the **`drslot`** command manages a dynamically reconfigurable slot, such as a hot-plug slot. Use online manual pages to see specific usage and flags for the **`lsslot`** and **`drslot`** commands.

Method 3

This method is used when a disk is part of a Virtual I/O Server logical volume. Perform the following steps:

1. On the Virtual I/O Server, allocate a logical unit to the target client.
2. Run **`cfgmgr`** from the receiving client to configure the newly inserted disk.
3. Add the disk to a volume group for use using the **`extendvg`** command.

5.7.4 Disk replacement procedure

This section explains how and when to replace a physical volume. It is good practice to take a recent backup of your data before removing or replacing disks on the system. Consider the following questions before replacing a disk:

- Is the disk mirrored?
- Can it be replaced online without shutting down?
- Is it a hot spare?
- Is it part of rootvg?
- Can the data be salvaged and migrated?
- Is it part of the root volume group?

Replacing a mirrored disk

A mirrored disk can be replaced without losing data. This is because the second or third member disk has a copy of the data. We highly recommend that all critical systems have their data mirrored.

Follow this procedure to replace a mirrored disk:

1. Follow the steps to add a new disk in 5.7.3, “Adding a physical disk” on page 174. The new disk must have capacity at least as large as the failed disk.
2. Run the configuration manager to configure the new disk: **`cfgmgr`** or **`mkdev -l`**.
3. Replace the physical volume so that it can begin using the new disk, using the following **`replacepv`** ***`olddisk newdisk`*** command

Note: If the mirror for the logical volume is stale, the **replacepv** command does not work correctly.

The **replacepv** command is similar to running the **extendvg** command to introduce a disk to a volume group, the **migratepv** command to move data, and then the **reducevg** command to remove the old disk from the volume group.

Additional steps for a mirrored rootvg

For the rootvg volume group, you must also run the following commands to clear the boot image from the failed disk and add the new disk to the boot image:

```
chpv -c failingdisk
bootlist newdisk
bosboot -a -d /dev/newdisk
```

The **chpv -c** command clears the boot image from the failing disk. The **bootlist** command adds the new disk to the list of possible boot devices from which the system can be booted. The **bosboot -a** command creates a complete boot image on the default boot logical volume.

Replacing a non-mirrored disk

It is important to have regular backups of all non-mirrored disks because it is not always possible to salvage data in a situation where a disk is not mirrored. All vital data such as databases should always be mirrored.

If the data can be retrieved from the failing disk, perform the following steps:

1. Follow the steps to add a new disk drive in 5.7.3, “Adding a physical disk” on page 174.
2. Add the new disk to the volume group using the **extendvg** command.
3. Move the data from the failing disk to the new disk using the **migratepv** command.

   Note: Steps 2 and 3 can be replaced with the **replacepv** command.

4. Remove the failing disk from the system.

If the data is corrupted and cannot be salvaged, perform the following steps:

1. Remove the failing disk from the system.
2. Follow the steps to add a new disk drive.
3. Add the new disk to the volume group.

4. Restore the data from backup.

5.7.5 Volume group management

After adding a physical volume, the next step is to assign it to a volume group. Prior to AIX V4.3.1, standard volume groups were used. AIX Version 4.3.1 implemented a new volume group factor (also known as a t factor), which can be specified by the **`-t`** flag of the **`mkvg`** and the **`chvg`** commands. AIX Version 4.3.2 expanded the LVM scalability by introducing a big volume group, which is specified by the use of a **`-B`** flag. AIX 5L Version 5.3 further introduces scalable volume group (scalable VG). Scalable volume groups have the advantage of growth when needed. Be careful not to define a volume too large, because it will come with a reduction in performance because the internal tables require more time to search and other data areas are larger.

Table 5-2 shows the variation of configuration limits with the different VG types.

Table 5-2 Configuration limits for volume groups

VG type	Maximum PVs	Maximum LVs	Maximum PPs per VG	Maximum PP size
Normal VG	32	256	32512 (1016 * 32)	1 GB
Big VG	128	512	130048 (1016 * 128)	1 GB
Scalable VG	1024	4096	2097152	128 GB

The following user commands were changed in support for the new scalable volume group type: **`chvg`**, **`lsvg`**, and **`mkvg`**.

Notes:

- The conversion of a previously configured standard or big VG to a scalable VG type requires the volume group to be varied off (with the **`varyoffvg`** command).
- AFter the volume group is converted, it cannot be imported into AIX 5L Version 5.2 or previous versions.

The new scalable volume group is fully supported by the System Management Interface Tool (SMIT) and the Web-based System Manager graphical user interface. Example 5-16 on page 178 shows the added support for scalable volume groups.

Example 5-16 Added support for scalable volume groups

```
                              Add a Volume Group

Move cursor to desired item and press Enter.
Add an Original Volume Group
  Add a Big Volume Group
  Add a Scalable Volume Group


Esc+1=Help          Esc+2=Refresh       Esc+3=Cancel        Esc+8=Image
Esc+9=Shell         Esc+0=Exit          Enter=Do
```

Concurrent volume groups

Concurrent volume groups are used when several systems needs to access the same set of disks. This is typically used by HACMP and virtual shared disks.

> **Restriction:** HACMP and the virtual shared disk subsystem both use the concurrent LVM functions. To prevent conflicts, only one product can manage concurrent disks. A virtual shared disk server cannot be defined as part of a concurrent volume group cluster if HACMP is installed and is already managing concurrent volume groups. If HACMP is installed but not managing concurrent volume groups, concurrent virtual shared disks can be defined.

Root volume group

The root volume group is automatically created when installing the basic operating system or when restoring from mksysb. It is varied on automatically by the system using special procedures. The system is inaccessible without the root volume group. Wherever possible, administrators are advised not to place user data on the root volume group.

The root volume group contains:

- Startup files
- Base Operating System (BOS)
- System configuration information
- Optional software products

Non-root volume groups

This is a volume group typically used for user data. All file systems and logical volumes not used for running the system should be created in non-root volume groups. These are created as needed.

5.7.6 Logical volumes and file systems

Journaled file system (JFS) and enhanced journaled file system (JFS2) are supported by AIX 5L. If a 64-bit kernel is selected on BOS installation, the JFS2 support will also be enabled. JFS2 gives enhanced capabilities such as a maximum file size of 1 TB and maximum a file system size of 4 petabytes (PB).

> **Note:** Starting with AIX 5L Version 5.3, it is possible to shrink a file system if the new size is large enough to handle the data in the file system and the file system is created as an enhanced journaled file system (JFS2).

Within each volume group, one or more logical volumes (LVs) are defined. Logical volumes contain file systems, paging space, and raw data. Data on logical volumes appears to be contiguous to the user, but can be discontiguous on one or more physical volumes. If mirroring is specified for the logical volume, additional physical partitions are allocated to store the additional copies of each logical partition. Logical volumes are used either as a file system or as raw devices allocated for applications. AIX 5L creates the raw devices, but the management of the raw devices is up to the application. Applications place headers and trailers on the raw devices. For this reason, raw devices are normally backed up using a tool that plugs in to the application using the raw device. For example, backing up and restoring IBM DB2® when it uses raw devices requires a use of a tool such as IBM Tivoli Storage Manager. This subject is beyond the scope of this publication. Creating a logical volume requires thorough consideration.

Understanding the usage of the data in the logical volume is the first step to deciding how to create a logical volume. Here, we provide some guidelines (not rules) for creating logical volumes.

Consider the following items when creating logical volumes:

- Considerations for creating a logical volume for availability:
  - Using hot spares
  - Using mirroring
  - Using mirror write consistency check
- Considerations for creating a logical volume for performance:
  - Using multiple disks (*stripping*)
  - Using multiple adapters (*multipathing*)
  - Data placement on disk:
    - Edge

    - Middle
    - Center
    - Inner-middle
    - Inner-edge
    - Mirror Write Consistency Check (MWCC) on the outer edge
  - Allocation policy:
    - Normal
    - Strict
    - Super strict

5.8 Planning and performing backup and recovery

This section describe how to plan and perform AIX 5L backup and recovery. We also introduce some methods to back up non-root volume groups at the end of the section. These methods are only a subset of many other methods that can be used for non-root volume group backup and recovery.

5.8.1 System backup (mksysb)

This section describes how to create and verify a bootable system backup copy, a *mksysb* image of the root volume group. Before any system migration, upgrade, installation, or major system change, we recommend that you take a full backup of the system.

A mksysb is an image of the operating system with all fixes and configurations applied after the BOS installation. This backup can be used to reinstall a system to its original state after it has been corrupted. If the backup is created on tape, the tape is bootable and has the installation programs needed to install from the backup.You can use the Web-based System Manager, SMIT, or the **mksysb** command to make a backup image of the root volume group.

A system backup transfers the following configurations from the source system to the target system:

- Root volume group information
- Paging space information
- Logical volume information
- Placement of logical partitions (if creating map files has been selected in the Web-based System Manager or SMIT)

Note: We do not recommend the use of map files if you plan to reinstall the backup to target systems other than the source system, or when the disk configuration of the source system is to be changed before reinstalling the backup.

The Web-based System Manager or the SMIT backup menu lets you preserve configuration information, thus avoiding some of the configuring tasks normally required after restoring a system backup. A backup preserves the configuration if the following conditions are met:

- The target system has the same hardware configuration as the source system.
- The target disk has enough space to hold the backup image.

Both the Web-based System Manager and SMIT use the **mksysb** command to create a backup image, stored either on CD, DVD, tape, or in a file.

System backup to tape

This method creates a bootable operating system backup to tape.

The basic requirements for backing up to tape are:

- A tape drive must be assigned on both the source and the target partition.
- A tape must be present in the tape device and it must be in writable.
- There is enough space in /tmp or rootvg; the **-X** flag can be used to extend /tmp when needed.

To create a system backup, perform the following steps. It is assumed that /dev/rmt0 is the available tape device on the system.

1. Run **smitty mksysb**.
2. Press F4 or Esc+4 to get a list of available devices.
3. Select the required device and press Enter.

   Example 5-17 shows an example of **smitty mksysb**.

You can also use the **mksysb -i /dev/rmt0** command.

Example 5-17 Example of the smitty mksysb command

```
Back Up the System

Type or select values in entry fields.
Press Enter AFTER making all desired changes.

[TOP]                                                  [Entry Fields]
```

```
       WARNING:  Execution of the mksysb command will
                 result in the loss of all material
                 previously stored on the selected
                 output medium. This command backs
                 up only rootvg volume group.

* Backup DEVICE or FILE                                [/dev/rmt0]            +/
  Create MAP files?                                     no                    +
  EXCLUDE files?                                        no                    +
  List files as they are backed up?                     no                    +
  Verify readability if tape device?                    no                    +
  Generate new /image.data file?                        yes                   +
  EXPAND /tmp if needed?                                no                    +
Esc+1=Help          Esc+2=Refresh       Esc+3=Cancel        Esc+4=List
Esc+5=Reset         Esc+6=Command       Esc+7=Edit          Esc+8=Image
Esc+9=Shell         Esc+0=Exit          Enter=Do
```

Creating a system backup to file

This method creates a system backup to file. It is important to have enough space in the file system where the backup image is created. The backup image file must be excluded if the backup will be written to a rootvg file system. We do not recommend writing the image to rootvg. The file creates a system image, but the image cannot be used for boot or install operations. You need to boot or install from the product CD or NIM master to use this file for system recovery. It is possible to create a bootable CD/DVD from this image using the **mkcd** command for CD-R or the **mkdvd** command for DVR-R. Another way to use a backup image created to a file is by using NIM to create a mksysb Shared Product Object Tree (SPOT) resource.

To create the image, use one of the following methods:

- Run the **smitty mksysb** command and select the name of the file to contain the backup image from the Backup Device or File field. See Example 5-17 on page 181, which shows where to specify a file or device. Instead of a device, list the full path name of your backup image file.
- Run the **mksysb -i** ***/mymksysbfile*** command.

System backup to DVD-R/CD-R

The **mkcd** command creates a multivolume CD (or CDs) from a mksysb or savevg backup image. If used with the **-L** flag, the **mkcd** command can create the backup to a DVD-R device. You can also use the **mkdvd** command to create a bootable mksysb to DVD. You can elect to create a bootable CD/DVD from an existing image or create a new mksysb image using the **mkcd** or **mkdvd** commands. You are required to have a CD-R or DVD-R drive configured. When using the **mkcd -L** or **mkdvd** command to write to DVD, the system will expect the use of 4 GB before it writes to a multiple DVD. Avoid using a DVD-R smaller than 4 GB; otherwise, the backup will fail.

5.8.2 System recovery methods

This section describes how to restore the basic operating system (BOS) using a system backup image, the mksysb image. You can install a system from a backup image that is stored on tape, CD, or DVD, or in a file.

BOS reinstallation from system backup is useful for:

- Restoring a corrupted operating system.
- Using an existing system image to reduce (or even eliminate) repetitive installation and configuration tasks.
- Transferring many user configuration settings to the target system (a different machine on which you are installing the system backup).
- It is also used to reduce the size of rootvg file systems. This is valid only when using AIX 5L Version 5.2 or earlier and if your rootvg is not using JFS2.
- The procedure used to install from a system backup depends on whether you are installing on the system where the backup was created, or on another system which might be of a different system architecture.
- Cloning a system backup helps to install a system backup on a target machine to propagate a consistent operating system, optional software, and configuration settings.

5.8.3 Cloning a system backup

Beginning with AIX 5L Version 5.2, all device and kernel support is automatically installed during a Base Operating System installation. With this feature, the system backup copies all known device drivers shipped with the AIX 5L Version 5.3 product. That means the mksysb image can be used to install the mksysb on heterogeneous systems with devices different from the source machine without the use of the product media to boot the system.

The installed device and kernel support filesets are follows:

- devices.*
- bos.mp
- bos.mp64, if necessary

5.8.4 Alternate disk installation

Alternate disk installation lets you install the operating system while it is still up and running. Alternate mksysb installation involves installing a mksysb image that has already been created from a system onto an alternate disk of the same system or another system. The alternate disk or disks cannot contain a volume group. With all device support, the target system does not have to be the same architecture as the system where the backup was created.

The following file sets are required for alternate disk installation:

- bos.alt_disk_install.boot_images

  Must be installed for alternate disk mksysb installations.

- bos.alt_disk_install.rte

  Must be installed for rootvg cloning and alternate disk mksysb installations.

Alternate disk installation can be used in the following ways:

- Installing a mksysb image on another disk while the system is running
- Cloning the current running rootvg to an alternate disk
- Using the Network Installation Management (NIM) environment to perform a alternate disk migration installation of a NIM client

The advantages of using alternate disk install include:

- Installing BOS on LPARs on the same system where NIM is unavailable, there is no CD or tape device, and Virtual I/O Server cannot be set up. This is done by installing a single image to many disks and creating preinstalled partitions using the disks.
- A system can be updated to another version on the alternate disk while the current disk is running. The upgrade or migration will only involve changing the boot list and restarting the system. This reduces the installation and upgrade down time considerably.
- Falling back to the old version after a migration or installation using alternate disk involves changing the bootlist and restarting the system.

The following steps show how to run the **alt_disk_install** command with some common flags:

1. Run the **alt_disk_install** command to begin cloning the rootvg on hdisk0 to hdisk1:

   ```
   # /usr/sbin/alt_disk_install -O -B -C hdisk1
   ```

   The cloned disk (hdisk1) will be named altinst_rootvg by default.

2. Rename the cloned disk (hdisk1) to alt1 so that you can repeat the operation with another disk:

```
# /usr/sbin/alt_disk_install -v alt1 hdisk1
```

3. Run the **alt_disk_install** command again to clone to another disk and rename the cloned disk, as follows:

```
# /usr/sbin/alt_disk_install -O -B -C hdisk2
# /usr/sbin/alt_disk_install -v alt2 hdisk2
```

Functional changes implemented in AIX 5L Version 5.3 include:

- The **alt_disk_install** command has been partitioned into separate modules with separate syntax based on operation and function.
- The man pages have been improved by creating a separate man page for each module.
- The following three new commands have been added:
  - **alt_disk_copy** creates copies of rootvg on an alternate set of disks.
  - **alt_disk_mksysb** installs an existing mksysb on an alternate set of disks.
  - **alt_rootvg_op** performs Wake, Sleep, and Customize operations.

> **Note:** The **alt_disk_install** module will continue to ship as a wrapper to the new modules. However, it will not support any new functions, flags, or features.

5.8.5 NIM client (mksysb installation)

A mksysb installation restores the BOS and additional software to a target from a mksysb image in the NIM environment.

The major prerequisites include:

- A mksysb image must be created and made available on the hard disk of the NIM master or a running NIM client.
- The NIM master must be configured. SPOT and mksysb resources must be defined. The mksysb resource can be created from the image in the previous prerequisite.
- The NIM client must be installed.
- There must be network connectivity between the target client and the NIM master. Performing a ping test from the client’s SMS menu can assist in confirming the network connectivity.
- The SPOT and mksysb resources should be at the same level of AIX when used for NIM BOS installations.

Note: We recommend that you either have no sparse files on your system or that you ensure that you have enough free space in the file system for future allocation of the blocks.

For more information about AIX 5L backup and recovery procedures, refer to the AIX information center, available at:

http://publib.boulder.ibm.com/infocenter/pseries/index.jsp

5.8.6 Using NIM with alternate disk install

The **alt_disk_install** command (available in AIX 4.3 or later) can be used to install a mksysb image on a NIM client's alternate disk or disks, or it can be used to clone a client running rootvg to an alternate disk. The target of an **alt_disk_install** operation can be a stand-alone NIM client or a group of stand-alone NIM clients. The clients must also have the bos.alt_disk_install.rte fileset installed.

The syntax for the **alt_disk_install** NIM mksysb operation is as follows:

```
nim -o alt_disk_install -a source=mksysb -a mksysb=mksysb_resource -a
disk=target_disk(s) -a attribute=Value.... TargetName|TargetNames
```

5.8.7 Using tape or CD/DVD-R

Restoring of a system using tape or CD/DVD requires that the device is installed on the system. No network connectivity is required to a NIM master. This is not always the preferred method because most system partitions do not have a tape or CD device attached. SP systems used nodes without these devices and also required network installation and booting.

Use the following basic steps, assuming that the required devices are configured:

1. Insert the disk or tape into the drive.
2. Remove all media in other drives that might have a BOS or mksysb image.
3. Switch the system off, and then on.
4. Press the appropriate key when the icons appear to change the bootlist so that the system boots from CD (5 or F5).
5. Change the bootlist and follow the BOS installation.
6. When BOS Welcome menu opens, select the **system recovery** option to start.
7. Select the option to **install from system backup**.

5.8.8 Non-root volume group backup and recovery

A user volume group, also called the *non-root* volume group, typically contains data files and application software.

Back up and restore of data on virtual disks

Data on virtual disks can be backed up and restored as usual from the client operating system. File system or database utilities can be used to accomplish these tasks, because the virtual disks behave exactly like physical disk drives. The advantage of a virtual disk is that it can be sized to match the requirements of the operating system. For example, a boot disk does not have to be a complete physical disk drive, especially because the drives designed currently can be larger than 128 GB, while the operating system uses less than 3 GB. The rest of the disk is wasted. Using virtual boot disk, you can create many systems on a 128 GB hdisk.

5.8.9 The savevg command

The **savevg** command finds and backs up all files belonging to a specified volume group. The volume group must be varied-on, and the file systems must be mounted. The **savevg** command uses the data file created by the **mkvgdata** command.

Restrictions:

- The **savevg** command will not save a file system that is not mounted.
- Although it will create a map of logical volumes created, it will not save the data in the logical volumes (raw device). Application tools such as IBM Tivoli Storage Manager are required to back up the data in the raw devices.

Other backup and restore commands include:

tar The **tar** command manipulates archives by writing files to, or retrieving files from, an archive storage medium.

cpio The **cpio** command copies files into and out of archive storage and directories.

5.8.10 Disaster recovery (business continuity)

It is important to understand the importance of data availability for critical business applications. Disaster recovery is a process that assists in bringing a business back to operation after incidental outages. Natural disasters such as earthquakes and floods and disasters such as blackouts and terrorism can affect entire neighborhoods and cities, disrupting your ability to deliver application services for days, weeks, or even months. Local clustering provides valuable protection from small-scale network, hardware, or software glitches. Classic methods of disaster recovery involved moving tapes off-site and recovery took too long. Businesses today seek protection from problems that reach beyond a single data center.

Costs play an important role in the disaster recovery plan. It is important for business to understand the cost of an outage and to align it with the investment made on their business continuity. IBM offers business continuity solutions including HACMP.

Tape and off-site disaster recovery method

It is possible to reduce costs by implementing a disaster recovery plan using the method of moving tapes off-site. This option must be thoroughly investigated because the cost of losing business at the time of recovery might be well more than the costs saved. The worst disadvantages of using an off-site tape disaster recovery plan is the time it takes to recover. It can take hours and even days.

This plan requires a good tape management solution (such as IBM Tivoli Storage Manager).

Use the following steps to perform off-site backup:

1. Back up the data to tape.
2. Get a list of tapes to be moved off-site.
3. Get a courier to move the tapes.
4. Move the tapes to the off-site location.

Use the following steps to recover using off-site tape management:

1. Move to disaster recovery site.
2. Prepare the hardware and operating system.
3. Prepare the tape management system.
4. Request the tapes from off-site.
5. Enter the tapes.
6. Start recovery.

Note:

- These steps can be taken while business is unavailable due to data loss.
- The transportation of tapes can be delayed by the environment, for example, traffic, accidents, and strikes.
- The actual restore, excluding manual intervention, will take about twice the amount of time it took to back up.
- This method cannot be used for 24x7 data centers. A high availability business continuity solution is required for 24x7 applications.

High Availability Cluster Multi-Processing

High Availability Cluster Multi-Processing (HACMP) has the following major features:

- Helps reduce unplanned outages and improve system availability.
- Offers ease of use through configuration wizards, auto-discovery, and a Web-based interface.
- Backup systems can be located at a remote site for geographic disaster recovery.
- Provides on demand failover to enable system maintenance without service interruption.

HACMP for AIX 5L helps businesses ensure availability with reliable system monitoring and failure response.

The HACMP Extended Distance (HACMP/XD) option extends the protection of HACMP for AIX to geographically remote sites to help ensure business continuity even if an entire site is disabled by catastrophe. HACMP/XD automatically manages the replication and synchronization of your Web site, databases, and other critical data and applications to a separate location from your primary operations and keeps this replicated data updated in real time.

Table 5-3 on page 190 lists major components of the HACMP and HACMP/XD methods of protection. Note that off-site mirroring is included; however, basic off-site mirroring can be accomplished without HACMP through the use of storage area network (SAN) storage.

Table 5-3 Backup feature comparison

Overview of high availability and data backup/recovery options			
Offering	**Location of backup data**	**Technology**	**Max. distance primary/backup processors/DASD**
HACMP	AIX/LVM mirror	SAN	15 kilometers; can be extended with DWDM[a] or CWDM[b]
HACMP with XD feature			
HACMP/XD with IBM ESS Metro-Mirror (formerly PPRC)	Remote ESS Metro-Mirror (PPRC) unit	ESCON® Fibre Channel	103 kilometers (> 103 km by special order latency impact over 40 km)
		Fibre Channel	300 km
	eRCMF	ESCON or Fibre Channel	> 103 km
HACMP/XD IP GLVM HAGEO	Any remote site	IP-based	Unlimited distance; must consider latency > 40 km

[a] DWDM: Dense Wave Division Multiplexor
[b] CWDM: Coarse Wave Division Multiplexor

Tivoli Storage Manager

Another backup and recovery product is IBM Tivoli Storage Manager. The Tivoli Storage Manager product set is an enterprise-wide solution integrating automated network backup, archive and restore, storage management, and disaster recovery. The Tivoli Storage Manager product set is ideal for heterogeneous, data-intensive environments, supporting more than 35 platforms and more than 250 storage devices across LANs, WANs, and SANs, plus providing protection for leading databases and e-mail applications.

6

Advanced POWER Virtualization

This chapter discusses the Virtualization Engine technologies that are now integrated into the IBM @server p5 servers, AIX 5L Version 5.3, and Linux.

The Advanced POWER Virtualization feature allows more flexibility in the use of the IBM @server p5 hardware. Virtualization can apply to microprocessors, memory, I/O devices, or storage. Fine-grain virtualization permits near instantaneous matching of workload to resources allocated, eschewing the wasted resources common to the one-server/one-application model of computing.

Note: Advanced POWER Virtualization is a no-charge feature (FC 7992) on the p5-590 and p5-595. On other @server p5 systems, it is a priced feature.

6.1 Overview of Advanced POWER Virtualization

The Advanced POWER Virtualization feature includes:

Micro-Partitioning Allows partitions to be created in units of less than one CPU. The processors on the system can be partitioned into as many as 10 LPARs per processor

Virtual networking A function of the POWER Hypervisor, virtual LAN allows secure communication between logical partitions without the need for a physical Ethernet adapter.

Virtual I/O Server Provides the capability for a single physical I/O adapter to be used by multiple logical partitions of the same server, allowing consolidation of I/O resources and minimizing the number of I/O adapters required.

Partition Load Manager (PLM)
Provides cross-partition workload management across the system LPARs.

> **Note:** Virtual Ethernet, Micro-Partitioning, LPAR, and dynamic LPAR are available without the Advanced POWER Virtualization feature when the server is attached to an HMC. Simultaneous multithreading (SMT) is available on the base hardware with no additional features required.

Figure 6-1 on page 193 shows how several of these technologies combine to provide you the flexibility to help meet your computing requirements.

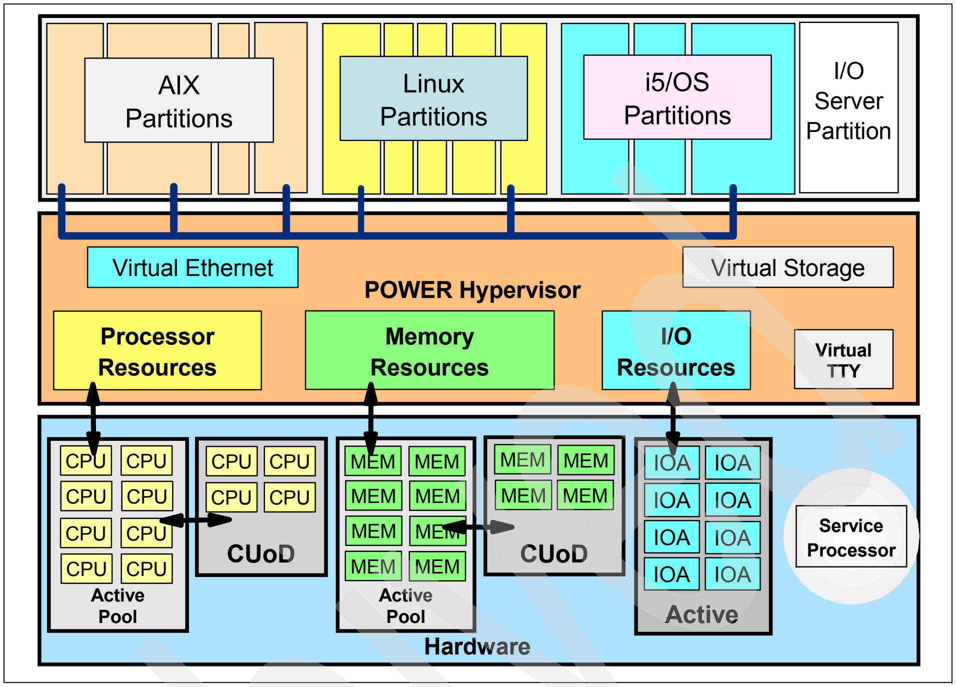


Figure 6-1 Virtualization technologies implemented on POWER5 servers

The Virtual I/O Server and Partition Load Manager (PLM) are licensed software components of the Advanced POWER Virtualization feature. They contain one charge unit per installed processor, including software maintenance. The initial software license charge for the Virtual I/O Server and PLM is included in the price of the Advanced POWER Virtualization feature. The related hardware features that include Virtual I/O Server and PLM are:

- 9117-570 FC 7942
- 9119-590 and 595 FC 7992

6.2 Micro-Partitioning

Micro-Partitioning is the ability to run more partitions on a server than there are physical processors by allocating fractions of processors to the partition. On POWER5 systems, you can choose between dedicated processor partitions and shared processor partitions using Micro-Partitioning.

The main benefit of Micro-Partitioning is that it allows increased overall utilization of system resources by automatically applying only the required amount of processor resource needed by each partition. The other advantage associated with this technology is that you can have up to 254 partitions running on a single, properly configured platform.

The POWER Hypervisor continually adjusts the amount of processor capacity allocated to each shared processor partition and any excess capacity unallocated based on current partition profiles within a shared pool. Tuning parameters enable the administrator extensive control over the amount of processor resources that each partition can use.

This section discusses the following topics about Micro-Partitioning:

- Shared processor partitions
- Shared pool overview

6.2.1 Shared processor partitions

Micro-Partitioning allows multiple partitions to share one physical processor. Partitions using Micro-Partitioning technology are referred to as shared processor partitions. Within this model, physical processors are abstracted into virtual processors that are then assigned to partitions, but the underlying physical processors are shared by these partitions. Virtual processor abstraction is implemented in the hardware and the POWER Hypervisor, a component of firmware. From an operating system perspective, a virtual processor is indistinguishable from a physical processor. The key benefit of implementing partitioning in the hardware is that it allows any operating system to run on POWER5 technology with little or no changes.

Partitions can be allocated in units as small as 1/10th of a processor and can be incremented in units of 1/100th of a processor. Each processor can be shared by up to 10 shared processor partitions. The shared processor partitions are dispatched and time sliced on the physical processors under the control of the POWER Hypervisor.

Table 6-1 shows the maximum number of logical partitions and shared processor partitions supported on the different models.

Table 6-1 Micro-Partitioning overview on p5 systems

p5 servers	Model 570	Model 590	Model 595
Processors	2-16	16-32	16-64
Dedicated processor partitions	2-16	16-32	16-64
Shared processor partitions	160	160-254	160-254

Shared processor partitions still need dedicated memory, but they do not more need physical I/O adapters using virtual Ethernet and virtual SCSI.

The shared processor partitions are created and managed by the HMC. When you start creating a partition, you have to choose between a shared processor partition and a dedicated processor partition.

When setting up a partition, you have to define the resources that belong to the partition, such as memory and I/O resources. For processor shared partitions, you need to configure these additional options:

- Minimum, desired, and maximum processing units of capacity
- The processing sharing mode, either capped or uncapped

  If the partition is uncapped, specify its variable capacity weight. An uncapped weight of 0 has the same effect as capped.
- Minimum, desired, and maximum virtual processors

We discuss these settings in the following sections.

Processing units of capacity

Processing capacity can be configured in fractions of 1/100th of a processor. The minimum amount of processing capacity that has to be assigned to a partition is 1/10 of a processor.

On the HMC, processing capacity is specified in terms of *processing units*. The minimum capacity of 1/10th of a processor is specified as 0.1 processing units. To assign a processing capacity representing 75% of a processor, 0.75 processing units are specified on the HMC.

On a system with two processors, a maximum of 2.0 processing units can be assigned to a partition. Processing units specified on the HMC are used to quantify the minimum, desired, and maximum amount of processing capacity for a partition.

After a partition is activated, processing capacity is usually referred to as capacity entitlement or entitled capacity.

Figure 6-2 on page 196 shows a graphical view of the definitions of processor capacity.

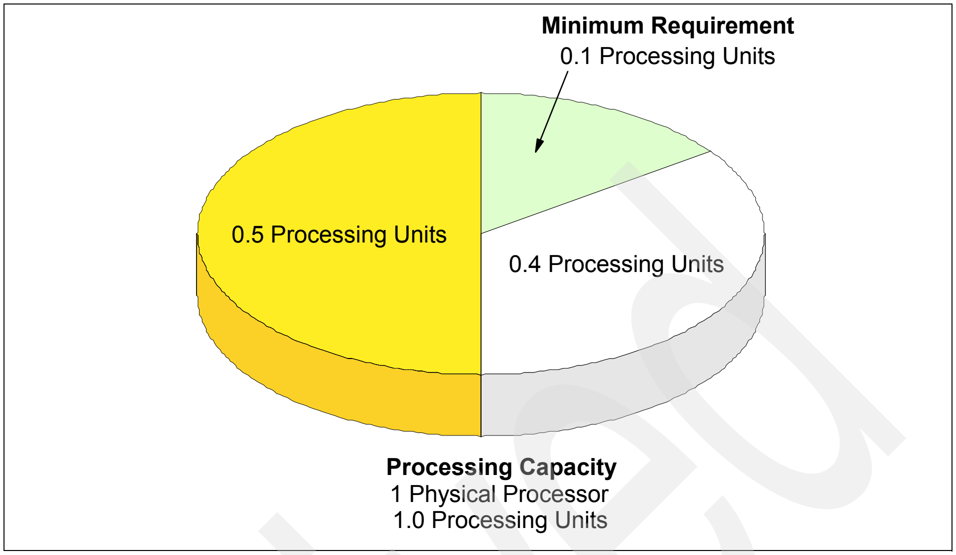


Figure 6-2 Processing units of capacity

Capped and uncapped mode

The next step in defining a shared processor partition is to specify whether the partition is running in a capped or uncapped mode:

Capped mode The processor unit never exceeds the assigned processing capacity.

Uncapped mode The processing capacity can be exceeded when the shared processing pools have available resources.

When a partition is run in an uncapped mode, you must specify the uncapped weight of that partition.

If multiple uncapped logical partitions require idle processing units, the managed system distributes idle processing units to the logical partitions in proportion to each logical partition's uncapped weight. The higher the uncapped weight of a logical partition, the more processing units the logical partition gets.

The uncapped weight must be a whole number from 0 to 255. The default uncapped weight for uncapped logical partitions is 128. A partition's share is computed by dividing its variable capacity weight by the sum of the variable capacity weights for all uncapped partitions. If you set the uncapped weight to 0, the managed system treats the logical partition as a capped logical partition. A logical partition with an uncapped weight of 0 cannot use more processing units than those that are committed to the logical partition. It is functionally identical to a capped partition.

Virtual processors

Virtual processors are the whole number of concurrent operations that the operating system can use. The processing power can be conceptualized as being spread equally across these virtual processors. Selecting the optimal number of virtual processors depends on the workload in the partition. Some partitions benefit from greater concurrence, while other partitions require greater power.

By default, the number of processing units that you specify is rounded up to the minimum number of virtual processors needed to satisfy the assigned number of processing units. The default settings maintain a balance of virtual processors to processor units. For example:

- If you specify 0.50 processing units, one virtual processor will be assigned.
- If you specify 2.25 processing units, three virtual processors will be assigned.

You also can use the Advanced tab in your partition profile to change the default configuration and to assign more virtual processors.

A logical partition in the shared processing pool will have at least as many virtual processors as its assigned processing capacity. By making the number of virtual processors too small, you limit the processing capacity of an uncapped partition. If you have a partition with 0.50 processing units and 1 virtual processor, the partition cannot exceed 1.00 processing units because it can only run one job at a time, which cannot exceed 1.00 processing units. However, if the same partition with 0.50 processing units was assigned two virtual processors and processing resources were available, the partition can use an additional 1.50 processing units.

6.2.2 Shared pool overview

The POWER Hypervisor schedules shared processor partitions from a set of physical processors that is called the shared processor pool. By definition, these processors are not associated with dedicated partitions.

In shared partitions, there is no fixed relationship between virtual processors and physical processors. The POWER Hypervisor can use any physical processor in the shared processor pool when it schedules the virtual processor. By default, it attempts to use the same physical processor, but this cannot always be guaranteed. The POWER Hypervisor uses the concept of a home node for virtual processors, enabling it to select the best available physical processor from a memory affinity perspective for the virtual processor that is to be scheduled.

Figure 6-3 on page 198 shows the relationship between two partitions using a shared processor pool of a single physical CPU. One partition has two virtual processors and the other a single one. The figure also shows how the capacity entitlement is evenly divided over the number of virtual processors.

When you set up a partition profile, you set up the desired, minimum, and maximum values you want for the profile. When a partition is started, the system chooses the partition's entitled processor capacity from this specified capacity range. The value that is chosen represents a commitment of capacity that is reserved for the partition. This capacity cannot be used to start another shared partition; otherwise, capacity might be overcommitted.

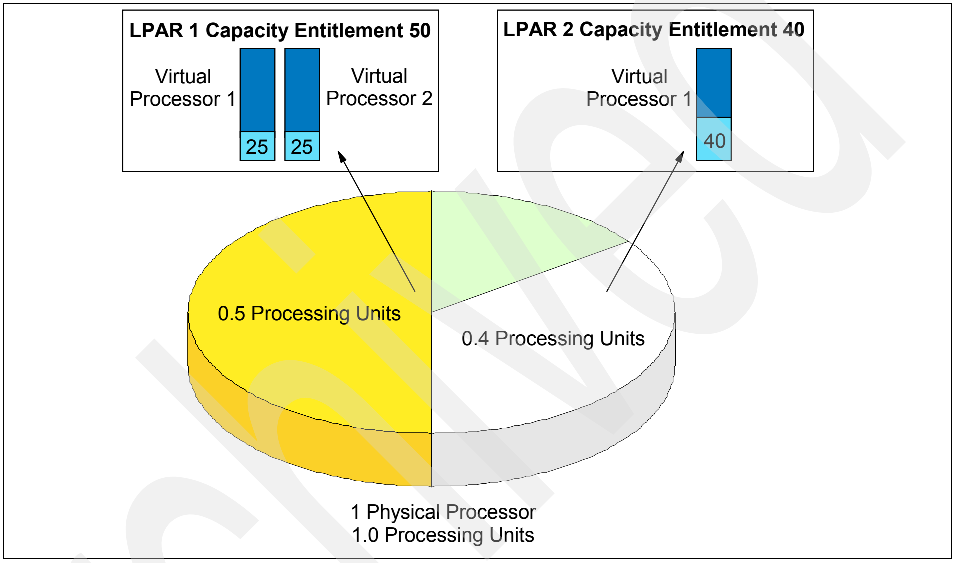


Figure 6-3 Distribution of capacity entitlement on virtual processors

When starting a partition, preference is given to the desired value, but this value cannot always be used because there might not be enough unassigned capacity in the system. In that case, a different value is chosen, which must be greater than or equal to the minimum capacity attribute. Otherwise, the partition cannot be started.

The entitled processor capacity is distributed to the partitions in the sequence in which the partitions are started. For example, consider a shared pool that has 2.0 processing units available.

Partitions 1, 2, and 3 are activated in sequence:

- Partition 1 activated
  Min. = 1.0, max. = 2.0, desired = 1.5
  Allocated capacity entitlement: 1.5

- Partition 2 activated
  Min. = 1.0, max. = 2.0, desired = 1.0
  Partition 2 does not start because the minimum capacity is not met.
- Partition 3 activated
  Min. = 0.1, max. = 1.0, desired = 0.8
  Allocated capacity entitlement: 0.5

The maximum value is only used as an upper limit for dynamic operations.

Figure 6-4 shows the usage of a capped partition of the shared processor pool. Partitions using the capped mode are not able to assign more processing capacity from the shared processor pool than the capacity entitlement will allow.

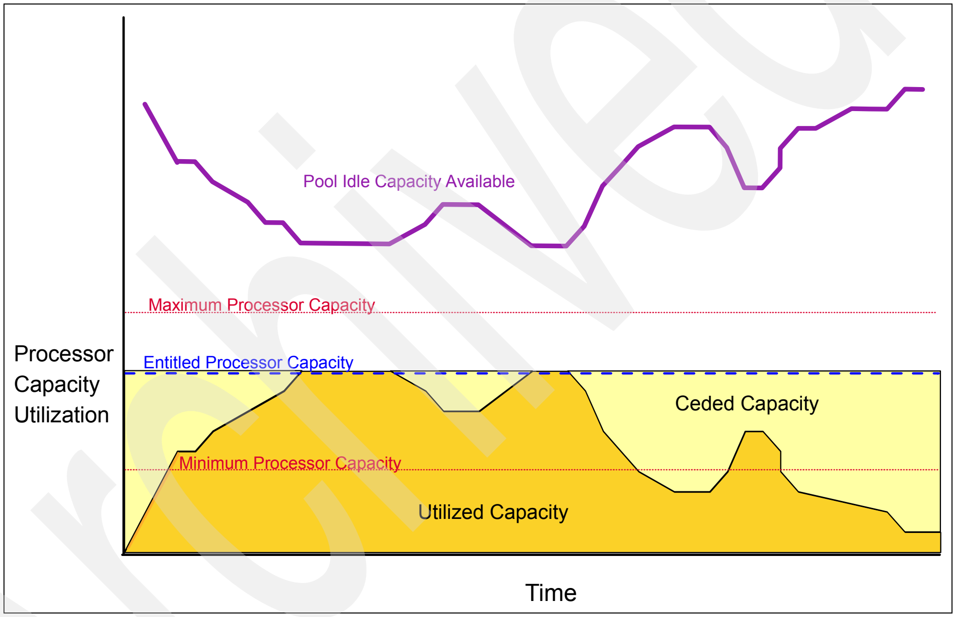


Figure 6-4 Capped shared processor partition

Figure 6-5 on page 200 shows the usage of the shared processor pool by an uncapped partition. The uncapped partition is able to assign idle processing capacity if it needs more than the entitled capacity.

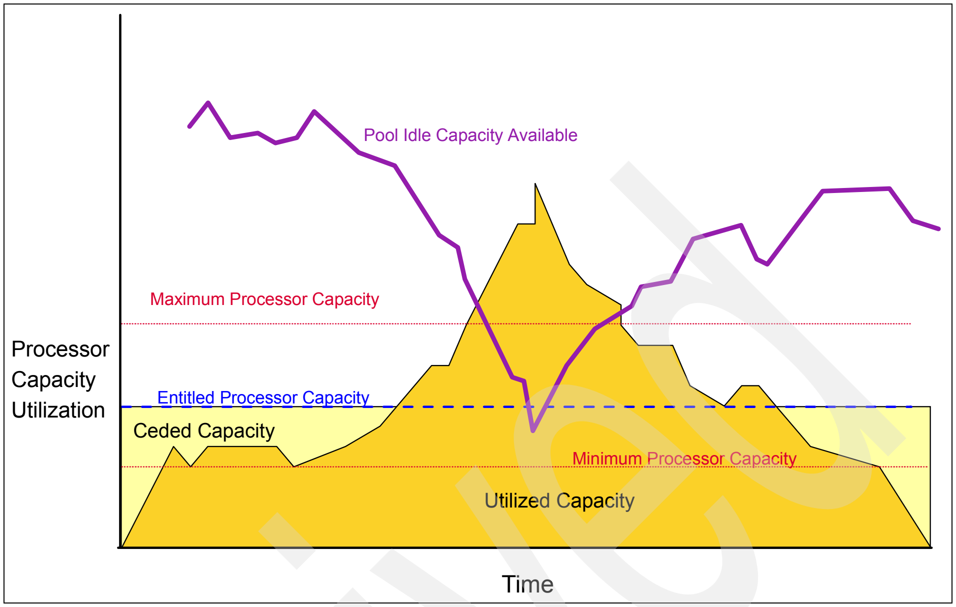


Figure 6-5 Uncapped shared processor partition

6.3 Virtual networking

Virtual Ethernet technology is supported on AIX 5L V5.3 on POWER5 hardware. This technology enables IP-based communication between logical partitions on the same system using a VLAN-capable software switch in POWER5 systems.

In this section, we discuss the following topics about virtual networking:

- Virtual LAN
- Virtual Ethernet
- Shared Ethernet Adapter

6.3.1 Virtual LAN (VLAN)

A virtual LAN is an internal LAN that connects a set of partitions. These LANs exist only in the memory of the server and are implemented through the POWER Hypervisor. They provide fast communication between partitions through virtual Ethernet and look like Ethernet LANs to the operating system in the partition.

This section discusses the concepts of virtual LAN (VLAN) technology with specific reference to its implementation within AIX 5L.

Virtual LAN overview

Virtual LAN is a technology used for establishing virtual network segments on top of physical switch devices. Typically, a VLAN is a broadcast domain that enables all nodes in the VLAN to communicate with each other without any L3 routing or inter-VLAN bridging. The use of VLAN provides increased LAN security and flexible network deployment over traditional network devices.

AIX 5L virtual LAN support

Some of the technologies for implementing VLANs include:

- IEEE 802.1Q VLAN
- Port-based VLAN
- Layer 2 VLAN

VLAN is described by the IEEE 802.1Q standard. VLAN is a method to logically segment a physical network such that Layer 2 connectivity is restricted to members that belong to the same VLAN. This separation is achieved by tagging Ethernet packets with their VLAN membership information and then restricting delivery to members of that VLAN.

The VLAN tag information is referred to as VLAN ID (VID). Ports on a switch are configured as being members of VLAN designated by the VID for that port. The default VID for a port is referred to as the port VID (PVID). The VID can be added to an Ethernet packet either by a VLAN aware host or by the switch in the case of VLAN unaware hosts.

A port will only accept untagged packets or packets with a VLAN ID (PVID or additional VIDs) tag of the VLANs to which the port belongs. A port configured in the untagged mode is only allowed to have a PVID and will receive untagged packets or packets tagged with the PVID. The untagged port feature helps systems that do not understand VLAN tagging communicate with other systems using standard Ethernet.

Each VLAN ID is associated with a separate Ethernet interface to the upper layers (for example, IP) and creates unique logical Ethernet adapter instances per VLAN (for example, ent1 or ent2). You can configure multiple VLAN logical devices on a single system.

6.3.2 Virtual Ethernet adapter

Virtual Ethernet enables interpartition communication without requiring any additional hardware. The POWER Hypervisor provides a virtual Ethernet switch function based on the IEEE 802.1Q VLAN standard that enables the administrator to define in-memory, point-to-point connections between partitions within the same server. These connections exhibit characteristics similar to physical high-bandwidth Ethernet connections and support multiple protocols (IPv4, IPv6, and ICMP).

The virtual Ethernet adapters can be dynamically created, and you can use the Hardware Management Console to make the VID assignments. The POWER Hypervisor transmits packets by copying the packet directly from the memory of the sender partition to the receive buffers of the receiver partition without any intermediate buffering of the packet.

Figure 6-6 is an example of an interpartition VLAN.

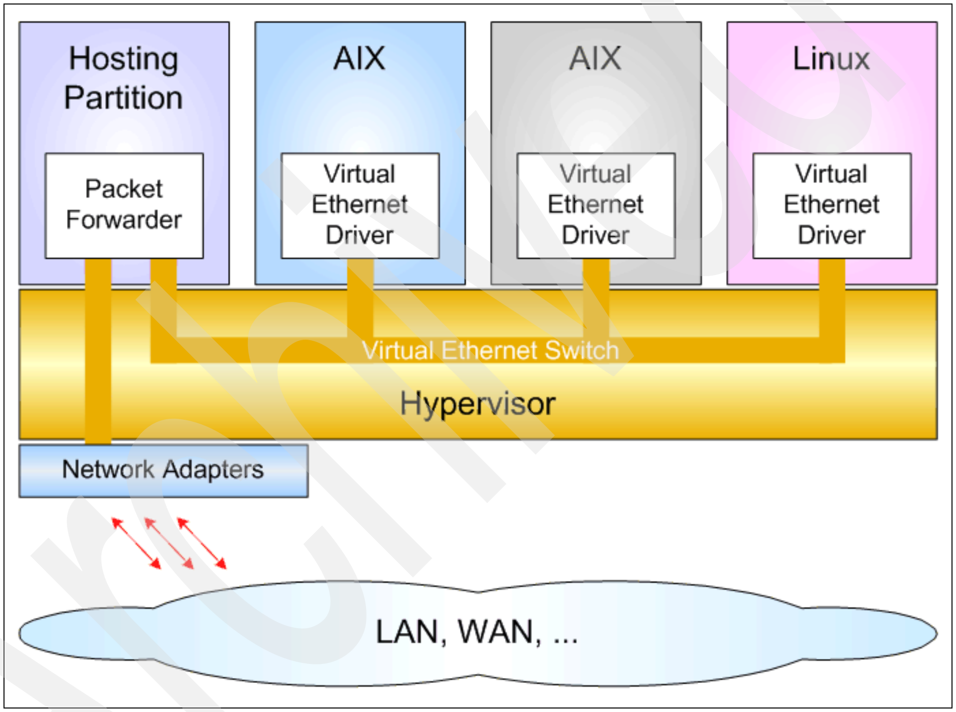


Figure 6-6 Logical view of an interpartition VLAN

6.3.3 Shared Ethernet Adapter (SEA)

The Shared Ethernet Adapter (SEA) hosted in the Virtual I/O Server acts as a Layer 2 switch between the internal and external VLAN using a physical network adapter.

The SEA must run in a Virtual I/O Server partition. The advantage of the SEA is that partitions can communicate outside the system without having a physical network adapter attached to the partition.

Shared Ethernet Adapters are configured in the Virtual I/O Server partition. Setting up a SEA requires one or more physical Ethernet adapters assigned to the I/O partition and one or more virtual Ethernet adapters with the trunk property defined using the HMC. The physical side of the SEA is either a single Ethernet adapter or a link aggregation of physical adapters. The link aggregation can also have an additional Ethernet adapter as a backup in case of failures in the network.

A single SEA setup can have up to 16 virtual Ethernet trunk adapters, and each virtual Ethernet trunk adapter can support up to 20 VLAN networks. Therefore, it is possible for a single physical Ethernet to be shared between 320 internal VLANs. The number of Shared Ethernet Adapters that can be set up in a Virtual I/O Server partition is limited only by the resource availability because there are no configuration limits.

6.3.4 Virtual networking example

This section shows how virtual networking can be implemented to allow communication between partitions and external networks in more detail.

The sample configuration in Figure 6-7 on page 204 uses four client partitions (Partition 1 through Partition 4) and one Virtual I/O Server. Each of the client partitions is defined with one virtual Ethernet adapter. The Virtual I/O Server has a Shared Ethernet Adapter that bridges traffic to the external network.

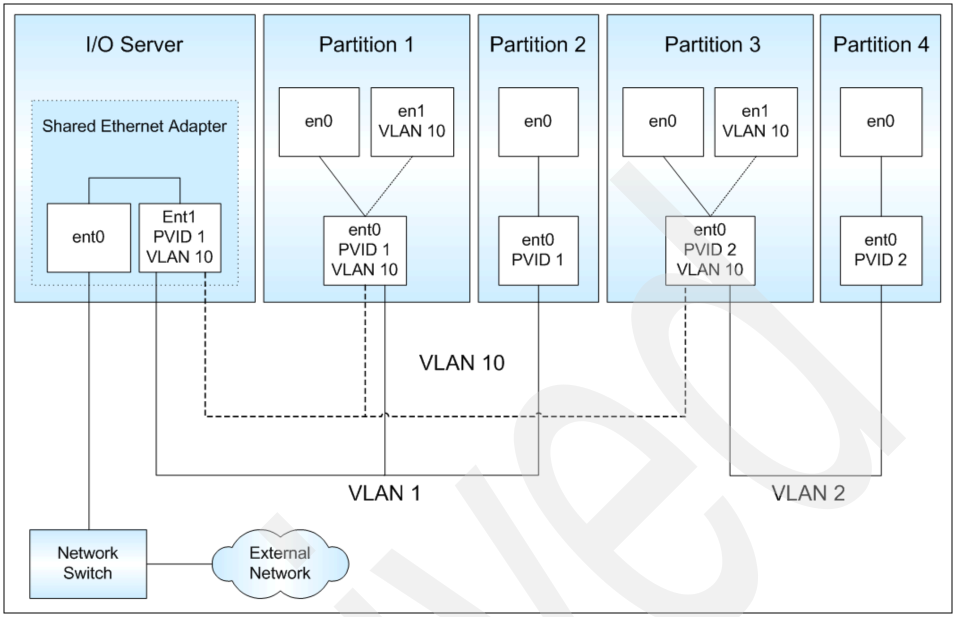


Figure 6-7 Virtual networking configuration

Interpartition communication

Partition 2 and Partition 4 use the port virtual LAN ID (PVID) only. This means that:

- Only packets for the VLAN specified as PVID are received.
- Packets sent have a VLAN tag added for the VLAN specified as PVID by the virtual Ethernet adapter.

In addition to the PVID, the virtual Ethernet adapters in Partition 1 and Partition 3 are also configured for VLAN 10 using a specific network interface (en1) created through **smitty vlan**. This means that:

- Packets sent through network interfaces en1 have a tag added for VLAN 10 by the network interface in AIX 5L.
- Only packets for VLAN 10 are received by the network interfaces en1.
- Packets sent through en0 are automatically tagged for the VLAN specified as PVID.
- Only packets for the VLAN specified as PVID are received by the network interfaces en0.

Table 6-2 lists which *client* partition can communicate which each other through what network interfaces.

Table 6-2 Interpartition VLAN communication

VLAN	Partition / network interface
1	Partition 1 / en0 Partition 2 / en0
2	Partition 3 / en0 Partition 4 / en0
10	Partition 1 / en1 Partition 3 / en1

Communication with external networks

The Shared Ethernet Adapter is configured with PVID 1 and VLAN 10. This means that untagged packets that are received by the Shared Ethernet Adapter are tagged for VLAN 1. Handling of outgoing traffic depends on the VLAN tag of the outgoing packets:

- Packets tagged with the VLAN that matches the PVID of the Shared Ethernet Adapter are untagged before being sent out to the external network.
- Packets tagged with a VLAN *other than* the PVID of the Shared Ethernet Adapter are sent out with the VLAN tag unmodified.

In our example, Partition 1 and Partition 2 have access to the external network through the network interface en0 using VLAN 1. Because these packets are using the PVID, the Shared Ethernet Adapter will remove the VLAN tags before sending the packets to the external network.

Partition 1 and Partition 3 have access to the external network using the network interface en1 and VLAN 10. These packets are sent out by the Shared Ethernet Adapter with the VLAN tag. Therefore, only VLAN-capable destination devices will be able to receive the packets. Table 6-3 lists this relationship.

Table 6-3 VLAN communication to external network

VLAN	Partition / Network interface
1	Partition 1 / en0 Partition 2 / en0
10	Partition 1 / en1 Partition 3 / en1

6.4 Virtual I/O Server

The IBM Virtual I/O Server is a special POWER5 partition that provides the ability to implement the virtual I/O function. Virtual I/O enables client partitions to share I/O resources.

We discuss the following topics related to the Virtual I/O Server:

- Overview of the Virtual I/O Server
- Virtual SCSI
- Installing the Virtual I/O Server
- Backup and maintenance
- Availability of the Virtual I/O Server partition

6.4.1 Overview of the Virtual I/O Server

The Virtual I/O Server is an appliance that provides virtual storage and shared Ethernet capability to client logical partitions on a POWER5 system. It allows a physical adapter on the Virtual I/O Server partition to be shared by one or more partitions, enabling clients to consolidate and potentially minimize the number of physical adapters.

The Virtual I/O Server is the link between the virtual and the real world. It can be seen as an AIX 5L-based appliance, and it is supported on POWER5 servers only. The Virtual I/O Server runs in a special partition that cannot be used for the execution of application code.

It mainly provides two functions:

- Server component for virtual SCSI devices (VSCI target)
- Support of Shared Ethernet Adapters for virtual Ethernet

Using the Virtual I/O Server facilitates the following functions:

- Sharing of physical resources between partitions on the system
- Creating partitions without requiring additional physical I/O resources
- Creating more partitions than there are I/O slots or physical devices available with the ability for partitions to have dedicated I/O, virtual I/O, or both
- Maximizing physical resource use on the system

The generic term virtual I/O relates to five different concepts:

- Three adapters: virtual SCSI, virtual Ethernet, and virtual Serial
- A special AIX 5L partition, called the Virtual I/O Server

- A mechanism to link virtual network devices to real devices, called Shared Ethernet Adapter (SEA)

Figure 6-8 shows an organization view of Micro-Partitioning including the Virtual I/O Server. The figure also includes virtual SCSI and Ethernet connections and mixed operating system partitions.

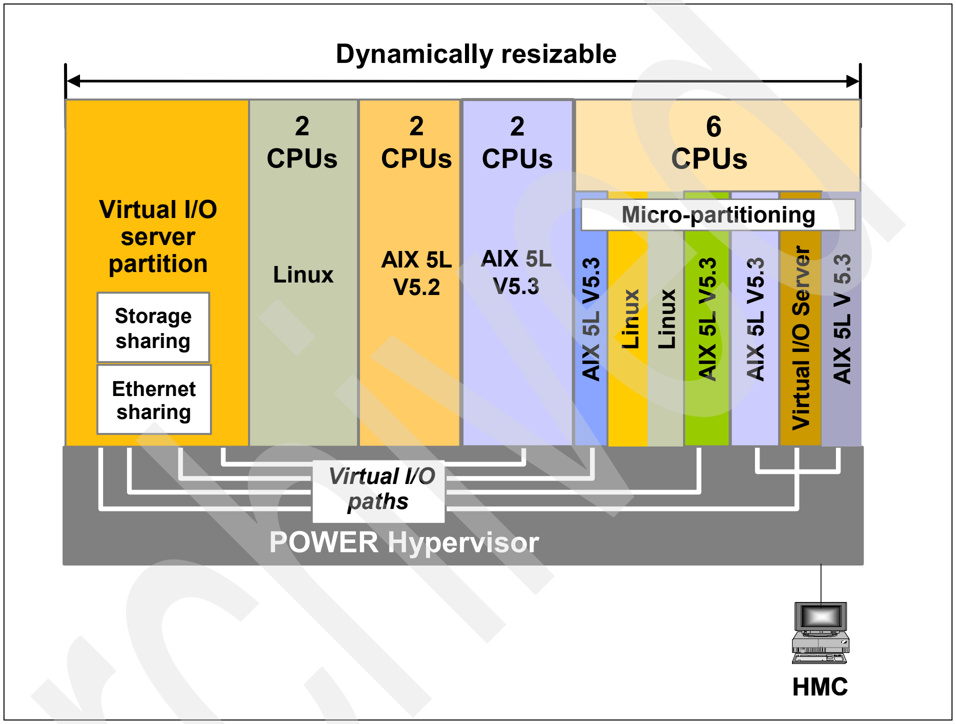


Figure 6-8 Virtual partition organization view

The Virtual I/O Server eliminates the requirement that every partition own a dedicated network adapter, disk adapter, and disk drive.

> **Note:** To increase the performance of I/O-intensive applications, dedicated physical adapters are preferred using dedicated partitions.

6.4.2 Virtual SCSI

Virtual SCSI is based on a client and server relationship. The Virtual I/O Server owns the physical resources and acts as a server or, in SCSI terms, a target device. The logical partitions access the virtual SCSI resources that are provided by the Virtual I/O Server as clients.

The virtual I/O adapters are configured using an HMC. The provisioning of virtual disk resources is provided by the Virtual I/O Server. Often, the Virtual I/O Server is also referred to as *hosting* partition, and the client partitions as *hosted* partitions.

Physical disks owned by the Virtual I/O Server can either be exported and assigned to a client partition as whole or can be partitioned into several logical volumes. The logical volumes can then be assigned to different partitions. Therefore, virtual SCSI enables sharing of adapters as well as disk devices.

To make a physical or a logical volume available to a client partition, it is assigned to a virtual SCSI server adapter in the Virtual I/O Server.

In Figure 6-9, one physical disk is partitioned into two logical volumes inside the Virtual I/O Server. Each of the two client partitions is assigned one logical volume, which it accesses through a virtual I/O adapter (virtual SCSI client adapter). Inside the partition the disk is seen as a normal hdisk.

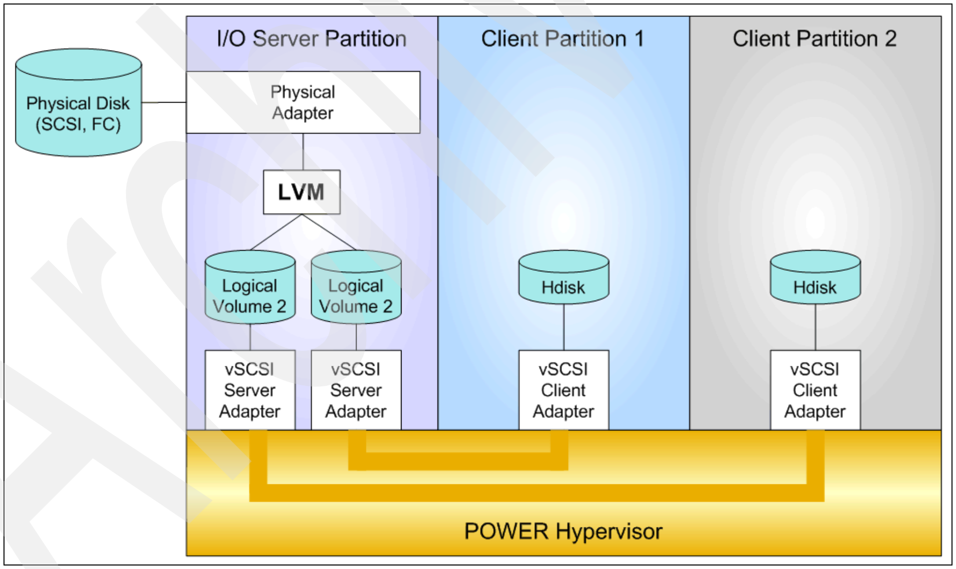


Figure 6-9 Virtual SCSI architecture overview

Virtual SCSI enables the attachment of previously unsupported storage solutions. As long as the Virtual I/O Server supports the attachment of a storage resource, any client partition can access this storage by using virtual SCSI adapters. For example, a Linux client partition can access an EMC storage attached to the Virtual I/O Server through a virtual SCSI adapter. Requests from the virtual adapters are mapped to the physical resources in the Virtual I/O Server. Therefore, driver support for the physical resources is needed only in the Virtual I/O Server.

For the latest list of Virtual I/O Server supported environments, refer to:

http://techsupport.services.ibm.com/server/vios/documentation/datasheet.html

6.4.3 Installing the Virtual I/O Server

You install the Virtual I/O Server partition from a special mksysb CD that is provided to clients that order the Advanced POWER Virtualization feature.

The Virtual I/O Server 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)
- NIM

The Virtual I/O Server is not accessible as a standard partition. Administrative access to the Virtual I/O Server partition is only possible as the user padmin, not as the root user. After login, the user padmin gets a restricted shell, which is not escapable, called the command line interface.

The Virtual I/O Server supports the following operating systems as virtual I/O clients:

- IBM AIX 5L Version 5.3
- SUSE LINUX Enterprise Server 9 for POWER
- Red Hat Enterprise Linux AS for POWER Version 3

6.4.4 Backup and maintenance

The following sections describe the backup and maintenance of the Virtual I/O Server.

Backup

Back up the Virtual I/O Server regularly using the **backupios** command to create an installable image of the root volume group onto either a bootable tape or a multi-volume CD/DVD. The creation of an installable NIM image on a file system is provided as well.

Additionally, the system’s partition configuration information, including the virtual I/O devices, should be backed up on the HMC. The client data should be backed up from the client system to ensure the consistency of the data.

The **backupios** command supports the following backup devices:

- Tape
- File system
- CD
- DVD

Maintenance

The Virtual I/O Server should be regarded as an appliance running in a dedicated partition. Fixes or upgrades for the Virtual I/O Server will be grouped into special fix packs and distributed separately from AIX 5L or other IBM operating systems.

To determine the level of Virtual I/O Server installed, run the **ioslevel** command. To download the latest level, refer to:

http://techsupport.services.ibm.com/server/vios/download/home.html

To install a fix pack, make a backup of the Virtual I/O Server and use the **updateios** command.

6.4.5 Availability of the Virtual I/O Server partition

The Virtual I/O Server partition is a critical partition; it always needs to be available for clients partitions. To achieve higher availability, you must install a second Virtual I/O Server, which provides further virtual SCSI disks and an Ethernet connection.

Because the Virtual I/O Server is an operating system-based appliance, redundancy for physical devices attached to the Virtual I/O Server can be provided by using capabilities such as Multipath I/O and IEEE 802.3ad Link Aggregation.

At the power-on of the system, the Virtual I/O Server partition can be started before all the client partitions using a system profile in the HMC.

Using the HMC, you can create and activate often-used collections of predefined partition profiles. This list of predefined partition profiles is called a system profile. The system profile is an ordered list of partitions and the profile to be activated for each partition. The system profile is referred to when the whole system is powered off and back on.

6.5 Partition Load Manager

The Partition Load Manager (PLM) software is part of the Advanced POWER Virtualization feature and helps clients maximize the utilization of processor and memory resources of dynamic logical partitioning (DLPAR)-capable logical partitions running AIX 5L on pSeries servers.

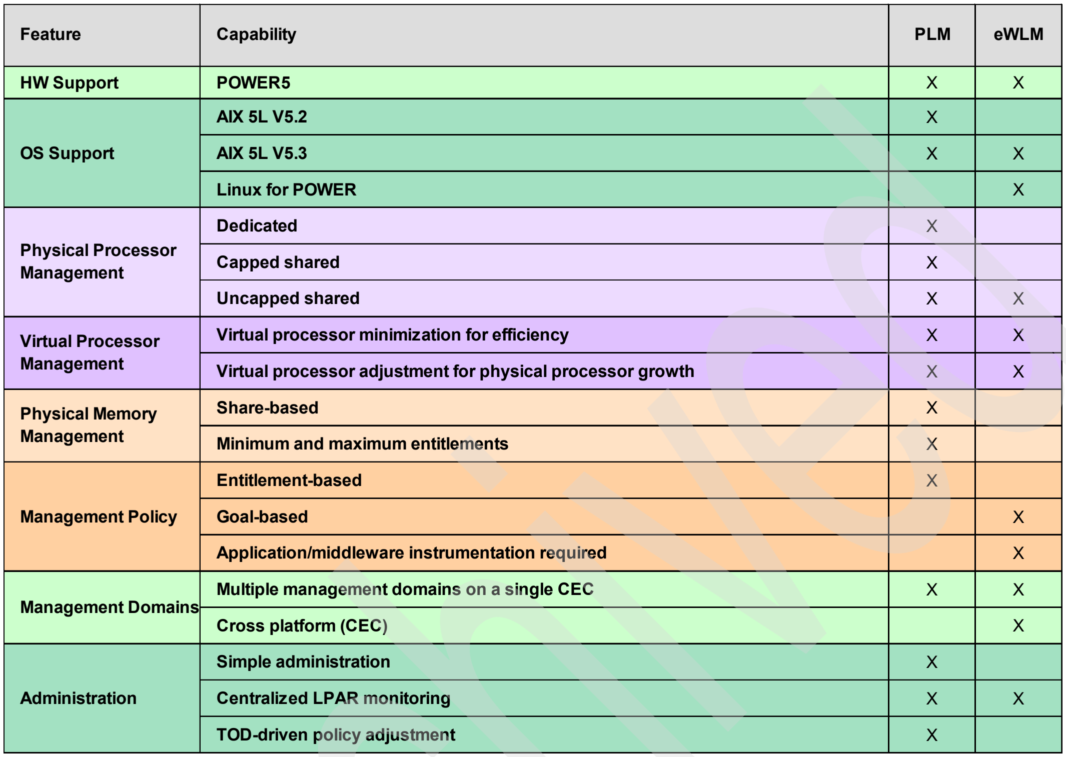
The Partition Load Manager is a resource manager that assigns and moves resources based on defined policies and utilization of the resources. PLM manages memory, both dedicated and shared processors, and partitions using Micro-Partitioning technology to readjust the resources. This adds additional flexibility on top of the Micro-Partitioning flexibility added by the POWER Hypervisor.

PLM, however, has no knowledge about the importance of a workload running in the partitions and cannot readjust priority based on the changes of the types of workloads; this is a function of another product, Enterprise Workload Manager, that can dynamically monitor and manage distributed heterogeneous workloads to achieve user-defined business goals.

Enterprise Workload Manager provides the following functions:

- End-to-end topology view and statistics for business transactions
- Goal-based resource optimization
- Physical resources improvement
- Workload management of heterogeneous environments
- End-to-end response time reporting

Figure 6-10 shows a comparison of features between the PLM and Enterprise Workload Manager.



Feature	Capability	PLM	eWLM
HW Support	POWER5	X	X
OS Support	AIX 5L V5.2	X	
	AIX 5L V5.3	X	X
	Linux for POWER		X
Physical Processor Management	Dedicated	X	
	Capped shared	X	
	Uncapped shared	X	X
Virtual Processor Management	Virtual processor minimization for efficiency	X	X
	Virtual processor adjustment for physical processor growth	X	X
Physical Memory Management	Share-based	X	
	Minimum and maximum entitlements	X	
Management Policy	Entitlement-based	X	
	Goal-based		X
	Application/middleware instrumentation required		X
Management Domains	Multiple management domains on a single CEC	X	X
	Cross platform (CEC)		X
Administration	Simple administration	X	
	Centralized LPAR monitoring	X	X
	TOD-driven policy adjustment	X	

Figure 6-10 Comparison of features of PLM and Enterprise Workload Manager

Partition Load Manager is set up in a partition or on another system running AIX 5L Version 5.2 ML4 or AIX 5L Version 5.3. A single instance of the Partition Load Manager can only manage a single server.

To configure Partition Load Manager, you can use the command line interface or the Web-based System Manager for a graphical setup.

Partition Load Manager uses a client/server model to report and manage resource utilization. The clients (managed partitions) notify the PLM server when resources are either under-utilized or over-utilized. Upon notification of one of these events, the PLM server makes resource allocation decisions based on a policy file defined by the administrator.

Partition Load Manager uses the Resource Monitoring and Control (RMC) subsystem for network communication, which provides a robust and stable framework for monitoring and managing resources. Communication with the

HMC to gather system information and execute commands PLM requires a configured SSH connection.

Figure 6-11 shows an overview of the components of Partition Load Manager.

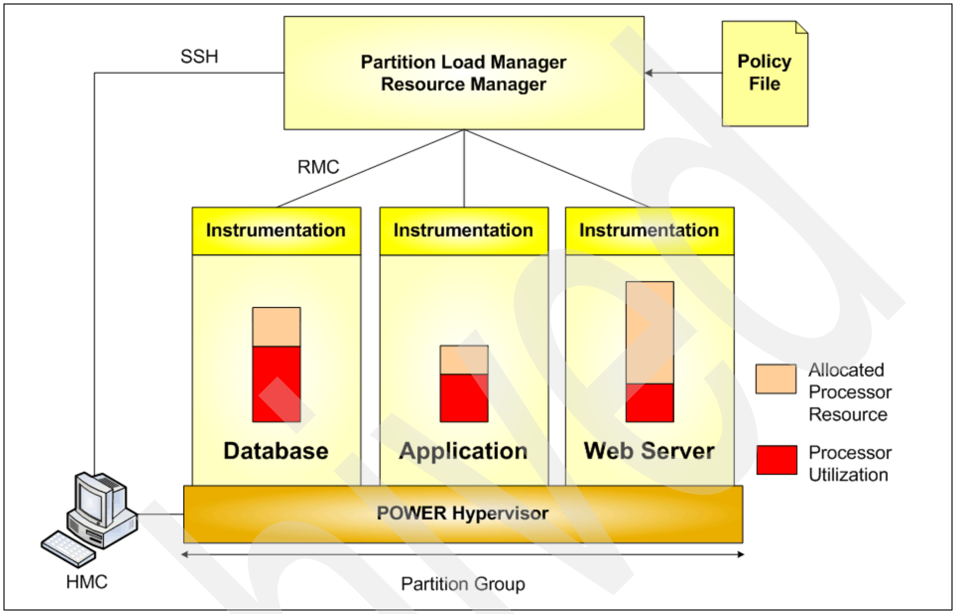


Figure 6-11 Partition Load Manager overview

The policy file defines managed partitions, their entitlements, and their thresholds and organizes the partitions into groups. Every node managed by the PLM must be defined in the policy file along with several associated attribute values:

- Optional maximum, minimum, and guaranteed resource values
- The relative priority or weight of the partition
- Upper and lower load thresholds for resource event notification

For each resource (processor and memory), the administrator specifies an upper and a lower threshold for which a resource event should be generated. You can also choose to manage only one resource.

Partitions that have reached an upper threshold become resource *requesters*. Partitions that have reached a lower threshold become resource *donors*. When a request for a resource is received, it is honored by taking resources from one of three sources when the requester has not reached its maximum value:

- A pool of free, unallocated resources

- A resource donor
- A lower priority partition with excess resources over the entitled amount

As long as there are resources available in the free pool, they will be given to the requester. If there are no resources in the free pool, the list of resource donors is checked. If there is a resource donor, the resource is moved from the donor to the requester. The amount of resource moved is the minimum of the delta values for the two partitions, as specified by the policy. If there are no resource donors, the list of excess users is checked.

When determining if resources can be taken from an excess user, the weight of the partition is determined to define the priority. Higher priority partitions can take resources from lower priority partitions. A partition's priority is defined as the ratio of its excess to its weight, where excess is expressed with the formula (current amount - desired amount) and weight is the policy-defined weight. A lower value for this ratio represents a higher priority. Figure 6-12 shows an overview of the process for partitions.

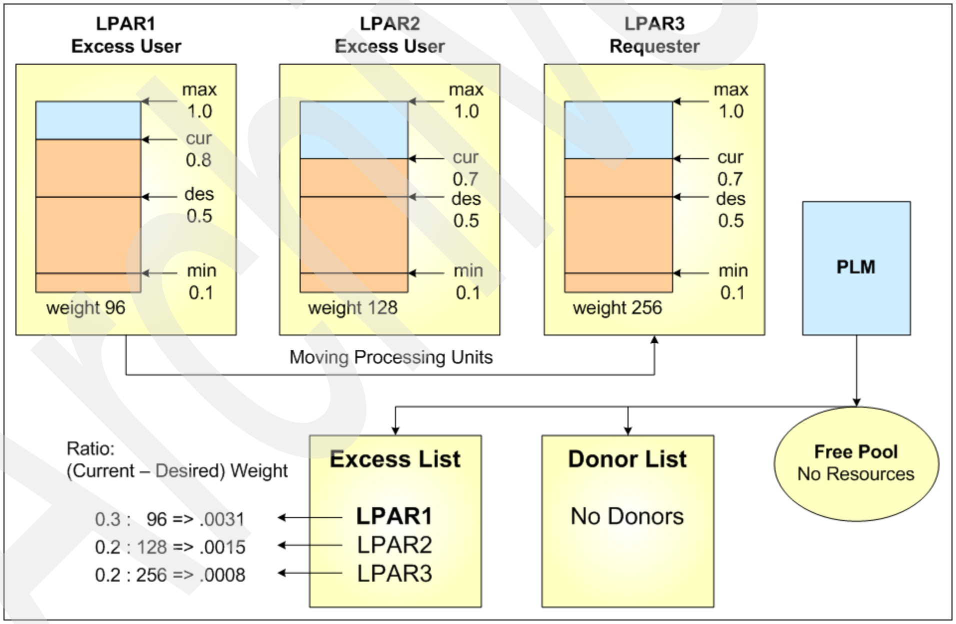


Figure 6-12 PLM resource distribution for partitions

In Figure 6-12, all partitions are capped partitions. LPAR3 is under heavy load and over its high CPU average threshold value, becoming a requestor. There are no free resources in the free pool and no donor partitions available. PLM now checks the excess list to find a partition having resources allocated over its guaranteed value and with a lower priority. Calculating the priority, LPAR1 has the highest ratio number and therefore the lowest priority. PLM deallocates resources from LPAR1 and allocates them to LPAR3.

If the request for a resource cannot be honored, it is queued and reevaluated when resources become available. A partition cannot fall below its minimum or rise above its maximum definition for each resource.

The policy file, when loaded, is static and has no knowledge of the nature of the workload on the managed partitions. A partition's priority does not change upon the arrival of high priority work. The priority of partitions can only be changed by some action, external to PLM, by loading a new policy.

Partition Load Manager handles memory and both types of processor partitions: dedicated and shared processor partitions. All the partitions in a group must be of the same processor type.

6.5.1 Memory management

PLM manages memory by moving logical memory blocks (LMBs) across partitions. To determine when there is demand for memory, PLM uses two metrics:

- Utilization percentage (ratio of memory in use to available)
- The page replacement rate

For workloads that result in significant file caching, the memory utilization on AIX 5L can never fall below the specified lower threshold. With this type of workload, a partition can never become a memory donor, even if the memory is not currently being used.

In the absence of memory donors, PLM can only take memory from excess users. Because the presence of memory donors cannot be guaranteed, and is unlikely with some workloads, memory management with PLM is only effective if there are excess users present. One way to ensure the presence of excess users is to assign each managed partition a low guaranteed value, such that it will always have more than its guaranteed amount. With this sort of policy, PLM will always be able to redistribute memory to partitions based on their demand and priority.

6.5.2 Processor management

For dedicated processor partitions, PLM moves physical processors, one at a time, from partitions that are not using them, to partitions that have a demand for them. This enables dedicated processor partitions running AIX 5L Version 5.2 and AIX 5L Version 5.3 to better use their resources. If one partition needs more processor capacity, PLM automatically moves processors from a partition that has idle capacity.

For shared processor partitions, PLM manages the entitled capacity and the number of virtual processors (VPs) for capped or uncapped partitions. When a partition has requested more processor capacity, PLM will increase the entitled capacity for the requesting partition if additional processor capacity is available. For uncapped partitions, PLM can increase the number of virtual processors to increase the partition's potential to consume processor resources under high load conditions. Conversely, PLM will also decrease entitled capacity and the number of virtual processors under low-load conditions to more efficiently use the underlying physical processors.

With the goal of maximizing a partition's and the system's ability to consume available processor resources, the administrator now has two choices:

- Configure partitions that have high workload peaks as uncapped partitions with a large number of virtual processors. This has the advantage of allowing these partitions to consume more processor resource when it is needed and available, with very low latency and no dynamic reconfiguration. For example, consider a 16-way system using two highly loaded partitions configured with eight virtual processors each, in which case, all physical processors could have been fully utilized. The disadvantage of this approach is that when these partitions are consuming at or below their desired capacity, there is a latency associated with the large number of virtual processors defined.
- Use PLM to vary the capacity and number of virtual processors for the partitions. This has the advantages of allowing partitions to consume all of the available processor resource on demand, and it maintains a more optimal number of VPs. The disadvantage to this approach is that, because PLM performs dynamic reconfiguration operations to shift capacity to and from partitions, there is a much higher latency for the reallocation of resources. Though this approach offers the potential to more fully use the available resource in some cases, it significantly increases the latency for redistribution of available capacity under a dynamic workload, because dynamic reconfiguration operations are required.

6.5.3 Limitations and considerations

You must consider the following limitations when managing your system with the Partition Load Manager:

- The Partition Load Manager can be used in partitions running AIX 5L Version 5.2 ML4 or AIX 5L Version 5.3.

- A single instance of the Partition Load Manager can only manage a single server. However, multiple instances of the PLM can be run on a single system, each managing a different server.
- The PLM cannot move I/O resources between partitions. Only processor and memory resources can be managed by the PLM.
- The PLM requires HMC Release 3 Version 2.6 or later on an HMC and an IBM eserver p5 system.

Abbreviations and acronyms

ABI	Application Binary Interface
AC	Alternating Current
ACL	Access Control List
AFPA	Adaptive Fast Path Architecture
AIO	Asynchronous I/O
APAR	Authorized Program Analysis Report
API	Application Programming Interface
ARP	Address Resolution Protocol
ASMI	Advanced System Management Interface
BFF	Backup File Format
BIND	Berkeley Internet Name Domain
BIST	Built-In Self-Test
BLV	Boot Logical Volume
BOOTP	Boot Protocol
BOS	Base Operating System
BSD	Berkeley Software Distribution
CA	Certificate Authority
CATE	Certified Advanced Technical Expert
CD	Compact Disk
CDE	Common Desktop Environment
CD-R	CD Recordable
CD-ROM	Compact Disk Read-Only Memory
CEC	Central Electronics Complex
CHRP	Common Hardware Reference Platform
CLI	Command Line Interface
CLVM	Concurrent LVM
CPU	Central Processing Unit
CRC	Cyclic Redundancy Check
CSM	Cluster Systems Management
CoD	Capacity on Demand
CUoD	Capacity Upgrade on Demand
DCM	Dual Chip Module
DES	Data Encryption Standard
DGD	Dead Gateway Detection
DHCP	Dynamic Host Configuration Protocol
DLPAR	Dynamic LPAR
DMA	Direct Memory Access
DNS	Domain Name System
DRM	Dynamic Reconfiguration Manager
DR	Dynamic Reconfiguration
DVD	Digital Versatile Disk
EC	EtherChannel
ECC	Error Checking and Correcting
EOF	End of File
EPOW	Early Power-Off Warning
ERRM	Event Response Resource Manager
ESS	Enterprise Storage Server®
F/C	Feature Code
FC	Fibre Channel
FCAL	Fibre Channel Arbitrated Loop
FDX	Full Duplex
FLOP	Floating Point Operation

FRU	Field Replaceable Unit
FTP	File Transfer Protocol
GDPS®	Geographically Dispersed Parallel Sysplex™
GID	Group ID
GPFS	General Parallel File System
GUI	Graphical User Interface
HACMP	High Availability Cluster Multi-Processing
HBA	Host Bus Adapter
HMC	Hardware Management Console
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol
Hz	Hertz
I/O	Input/Output
IBM	International Business Machines
ID	Identification
IDE	Integrated Device Electronics
IEEE	Institute of Electrical and Electronics Engineers
IP	Internetwork Protocol
IPAT	IP Address Takeover
IPL	Initial Program Load
IPMP	IP Multipathing
ISV	Independent Software Vendor
ITSO	International Technical Support Organization
IVM	Integrated Virtualization Manager
JFS	Journaled File System
JFS2	Enhanced Journaled File System
L1	Level 1
L2	Level 2
L3	Level 3
LA	Link Aggregation
LACP	Link Aggregation Control Protocol
LAN	Local Area Network
LDAP	Lightweight Directory Access Protocol
LED	Light Emitting Diode
LMB	Logical Memory Block
LPAR	Logical Partition
LPP	Licensed Program Product
LUN	Logical Unit Number
LV	Logical Volume
LVCB	Logical Volume Control Block
LVM	Logical Volume Manager
MAC	Media Access Control
Mbps	Megabits Per Second
MBps	Megabytes Per Second
MCM	Multichip Module
ML	Maintenance Level
MP	Multiprocessor
MPIO	Multipath I/O
MTU	Maximum Transmission Unit
NFS	Network File System
NIB	Network Interface Backup
NIM	Network Installation Management
NIMOL	NIM on Linux
NVRAM	Non-Volatile Random Access Memory
ODM	Object Data Manager
OSPF	Open Shortest Path First
PCI	Peripheral Component Interconnect
PIC	Pool Idle Count
PID	Process ID
PKI	Public Key Infrastructure
PLM	Partition Load Manager
POST	Power-On Self-Test

POWER	Performance Optimization with Enhanced RISC (Architecture)
PPC	Physical Processor Consumption
PPFC	Physical Processor Fraction Consumed
PTF	Program Temporary Fix
PTX®	Performance Toolbox
PURR	Processor Utilization Resource Register
PV	Physical Volume
PVID	Physical Volume Identifier
PVID	Port Virtual LAN Identifier
QoS	Quality of Service
RAID	Redundant Array of Independent Disks
RAM	Random Access Memory
RAS	Reliability, Availability, and Serviceability
RCP	Remote Copy
RDAC	Redundant Disk Array Controller
RIO	Remote I/O
RIP	Routing Information Protocol
RISC	Reduced Instruction Set Computer
RMC	Resource Monitoring and Control
RPC	Remote Procedure Call
RPL	Remote Program Loader
RPM	Red Hat Package Manager
RSA	Rivet, Shamir, Adelman
RSCT	Reliable Scalable Cluster Technology
RSH	Remote Shell
SAN	Storage Area Network
SCSI	Small Computer System Interface
SDD	Subsystem Device Driver
SMIT	System Management Interface Tool
SMP	Symmetric Multiprocessor
SMS	System Management Services
SMT	Simultaneous Multithreading
SP	Service Processor
SPOT	Shared Product Object Tree
SRC	System Resource Controller
SRN	Service Request Number
SSA	Serial Storage Architecture
SSH	Secure Shell
SSL	Secure Sockets Layer
SUID	Set User ID
SVC	SAN Virtualization Controller
TCP/IP	Transmission Control Protocol/Internet Protocol
UDF	Universal Disk Format
UDID	Universal Disk Identification
VIPA	Virtual IP Address
VG	Volume Group
VGDA	Volume Group Descriptor Area
VGSA	Volume Group Status Area
VLAN	Virtual Local Area Network
VP	Virtual Processor
VPD	Vital Product Data
VPN	Virtual Private Network
VRRP	Virtual Router Redundancy Protocol
VSD	Virtual Shared Disk
WLM	Workload Manager

Related publications

The publications listed in this section are considered particularly suitable for a more detailed discussion of the topics covered in this redbook.

IBM Redbooks

For information about ordering these publications, see “How to get IBM Redbooks” on page 226. Note that some of the documents referenced here may be available in softcopy only.

- *Linux Applications on pSeries*, SG24-6033
- *Managing AIX Server Farms*, SG24-6606
- *AIX Logical Volume Manager from A to Z: Introduction and Concepts*, SG24-5432
- *Effective System Management Using the IBM Hardware Management Console for pSeries*, SG24-7038
- *Introduction to pSeries Provisioning*, SG24-6389
- *i5/OS on ⓔserver p5 Models: A Guide to Planning, Implementation, and Operation*, SG24-8001
- *NIM: From A to Z in AIX 4.3,* SG24-5524
- *A Practical Guide for Resource Monitoring and Control (RMC)*, SG24-6615
- *Partitioning Implementations for IBM ⓔserver p5 Servers*, SG24-7039
- *Practical Guide for SAN with pSeries*, SG24-6050
- *Problem Solving and Troubleshooting in AIX 5L*, SG24-5496
- *Understanding IBM ⓔserver pSeries Performance and Sizing*, SG24-4810
- *Advanced POWER Virtualization on IBM System p5*, SG24-7940
- *IBM ⓔserver p5 570 Technical Overview and Introduction*, REDP-9117
- *IBM ⓔserver p5 590 and 595 Technical Overview and Introduction*, REDP-4024
- *IBM ⓔserver p5 590 and 595 System Handbook*, SG24-9119
- *AIX 5L Practical Performance Tools and Tuning Guide*, SG24-6478

Other publications

These publications are also relevant as further information sources.

The following types of documentation are available at the following URL:

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

- User guides
- System management guides
- Application programmer guides
- All commands reference volumes
- Files reference
- Technical reference volumes used by application programmers

Online resources

These Web sites and URLs are also relevant as further information sources:

- IBM Certification Web site

  http://www.ibm.com/certify
- AIX 5L operating system maintenance packages downloads

  http://www.ibm.com/servers/eserver/support/unixservers/aixfixes.html
- Autonomic computing on IBM @server pSeries servers

  http://www.ibm.com/autonomic/index.shtml
- Capacity on Demand

  http://www.ibm.com/servers/eserver/pseries/ondemand/cod/
- Hardware documentation

  http://publib16.boulder.ibm.com/pseries/en_US/infocenter/base/
- IBM @server Information Center

  http://publib.boulder.ibm.com/eserver/
- IBM @server pSeries and RS/6000 microcode update

  http://techsupport.services.ibm.com/server/mdownload2/download.html
- Support for AIX 5L and Linux servers

  http://www.ibm.com/servers/eserver/support/unixservers/index.html
- IBM @server support: Tips for AIX administrators

  http://techsupport.services.ibm.com/server/aix.srchBroker

- Hardware Management Console (HMC): Support for HMC for UNIX servers and Midrange servers

  `http://techsupport.services.ibm.com/server/hmc`

- Virtual I/O Server Supported Environment

  `http://techsupport.services.ibm.com/server/vios/documentation/datasheet.html`

- IBM LPAR Validation Tool for POWER processor-based systems

  `http://www.ibm.com/servers/eserver/iseries/lpar/systemdesign.html`

- IBM System p5, ⓔserver p5, pSeries, OpenPower and IBM RS/6000 Performance Report

  `http://www.ibm.com/servers/eserver/pseries/hardware/system_perf.html`

- **`nmon`** performance: A free tool to analyze AIX and Linux performance

  `http://www.ibm.com/developerworks/eserver/articles/analyze_aix/index.html`

- Service and productivity tools for Linux on POWER

  `http://techsupport.services.ibm.com/server/lopdiags`

- IBM Linux news: Subscribe to the Linux Line

  `https://www6.software.ibm.com/reg/linux/linuxline-i`

- Information about UnitedLinux for pSeries from Turbolinux

  `http://www.turbolinux.com`

- Linux on ⓔserver p5 and pSeries

  `http://www.ibm.com/servers/eserver/pseries/linux/`

- Microcode Discovery Service

  `http://techsupport.services.ibm.com/server/aix.invscoutMDS`

- POWER4 system microarchitecture, comprehensively described in the *IBM Journal of Research and Development*, Vol. 46, No.1, January 2002

  `http://www.research.ibm.com/journal/rd46-1.html`

- SCSI T10 Technical Committee

  `http://www.t10.org`

- SUSE LINUX Enterprise Server 8 for pSeries information

  `http://www.suse.de/us/business/products/server/sles/i_pseries.html`

- *Red Hat Enterprise Linux 3 Installation Guide for the IBM ⓔserver iSeries and IBM ⓔserver pSeries Architectures*

  `http://www.redhat.com/docs/manuals/enterprise/RHEL-3-Manual/ppc-multi-install-guide/`

- SUSE LINUX Enterprise Server 9 Documentation

  `http://www.novell.com/documentation/sles9/index.html`

How to get IBM Redbooks

You can search for, view, or download Redbooks, Redpapers, Hints and Tips, draft publications and Additional materials, as well as order hardcopy Redbooks or CD-ROMs, at this Web site:

ibm.com/redbooks

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services

Index

IBM @server p5 and pSeries Enterprise Technical Support AIX 5L V5.3

(0.5” spine)
0.475”<->0.875”
250 <-> 459 pages

IBM®

IBM @server Certification Study Guide @server p5 and pSeries Enterprise Technical Support AIX 5L V5.3

Redbooks

Developed specifically for the purpose of preparing for certification test 180

Makes an excellent companion to classroom education

For @server p5 and pSeries enterprise support professionals

This IBM Redbook is designed as a study guide for professionals wanting to prepare for the certification exam to achieve IBM Certified Systems Expert - @server p5 and pSeries Enterprise Technical Support AIX 5L V5.3. This technical support certification validates a broad scope of configuration, installation, and planning skills. In addition, it covers administrative and diagnostic activities needed to support logical partitions and virtual resources.

This publication helps IBM @server p5 and pSeries professionals seeking a comprehensive and task-oriented guide for developing the knowledge and skills required for the certification. It is designed to provide a combination of theory and practical experience needed for a general understanding of the subject matter.

This publication does not replace the practical experience you should have, but is an effective tool that, when combined with education activities and experience, should prove to be a very useful preparation guide for the exam. Due to the close association with the certification content, this publication might reflect older software and firmware levels of the IBM @server p5 systems and available features. If you are planning to take the @server p5 and pSeries Enterprise Technical Support AIX 5L V5.3 certification exam, this book is for you.

INTERNATIONAL TECHNICAL SUPPORT ORGANIZATION

BUILDING TECHNICAL INFORMATION BASED ON PRACTICAL EXPERIENCE

IBM Redbooks are developed by the IBM International Technical Support Organization. Experts from IBM, Customers and Partners from around the world create timely technical information based on realistic scenarios. Specific recommendations are provided to help you implement IT solutions more effectively in your environment.

For more information:
ibm.com/redbooks

SG24-7197-01 ISBN 0738494054