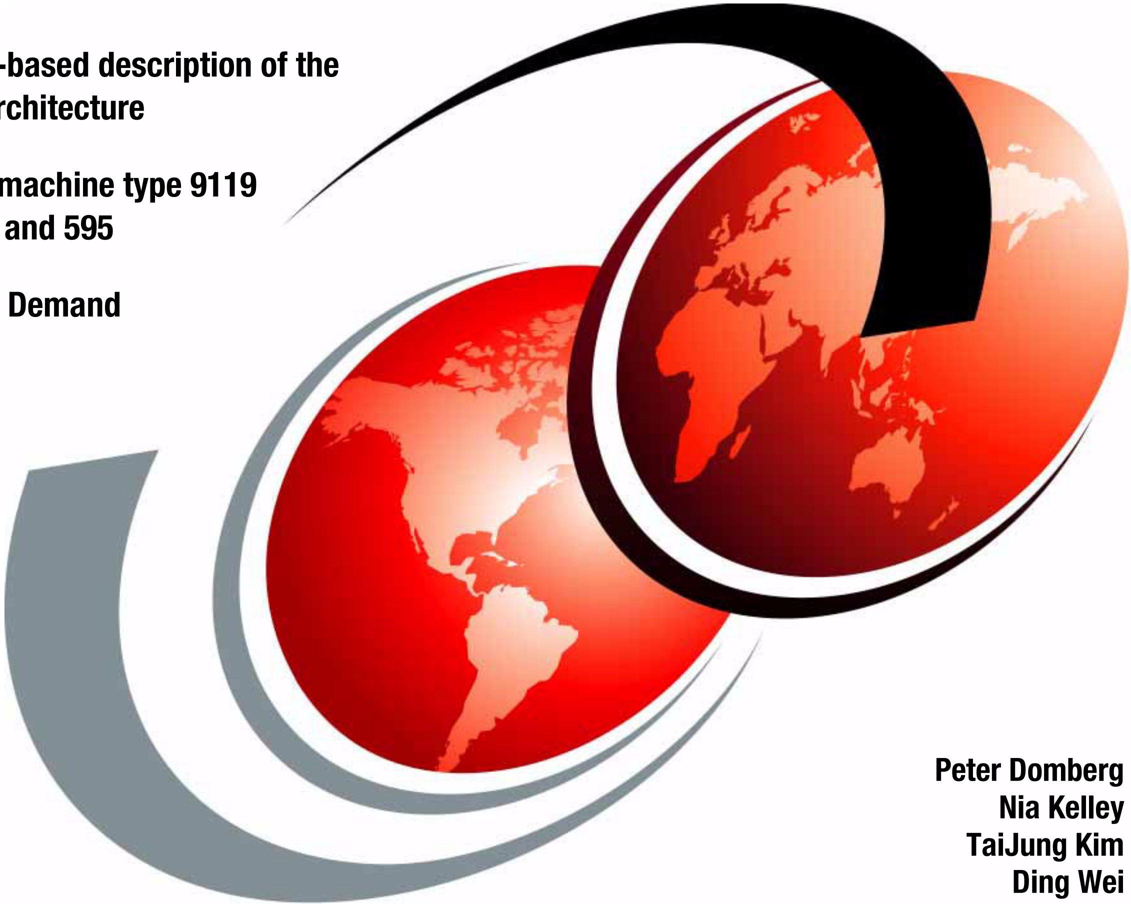
Draft Document for Review February 23, 2005 1:46 pm

SG24-9119-00

IBM eserver p5 590 and p5 595 System Handbook



Component-based description of the hardware architecture

A guide for machine type 9119 models 590 and 595

Capacity on Demand explained

Peter Domberg
Nia Kelley
TaiJung Kim
Ding Wei

Redbooks

ibm.com/redbooks

International Technical Support Organization

IBM eserver p5 590 and p5 595 System Handbook

February 2005

SG24-9119-00

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

First Edition (February 2005)

This edition applies to the IBM eserver p5 9119 Models 590 and 595.

Contents

Figures

Tables

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:

400®
BladeCenter™
Chipkill™
Electronic Service Agent™
Enterprise Storage Server®
ESCON®
@server®
@server®
Extreme Blue™
HACMP™
Hypervisor™
IBM®
Micro Channel®
Micro-Partitioning™
OpenPower™
Power Architecture™
Power PC®
POWER2™
POWER4+™
POWER4™
POWER5™
PowerPC®
POWER™
PS/2®
PTX®
Redbooks (logo) ™Balance®
Redbooks™
RS/6000®
S/390®
TotalStorage®
Versatile Storage Server™
Virtualization Engine™

The following terms are trademarks of other companies:

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

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

Intel, Intel Inside (logos), MMX, and Pentium are trademarks of Intel 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, and service names may be trademarks or service marks of others.

Preface

This IBM Redbook explores the IBM @server p5 models 590 and 595 (9119-590, 9119-595), a new level of UNIX servers providing world-class performance, availability, and flexibility. Ideal for on demand computing environments, data center implementation, application service providers, and high performance computing, this new class of high-end servers include mainframe-inspired self-management and security designed to meet your most demanding needs. The IBM @server p5 590 and p5 595 provide an expandable, high-end enterprise solution for managing the computing requirements needed to become an on demand business.

This publication includes the following topics:

- p5-590 and p5-595 overview
- p5-590 and p5-595 hardware architecture
- Virtualization features overview
- Capacity on Demand overview
- Reliability, availability, and serviceability overview
- Hardware Management Console features and functions

This publication is an ideal desk-side reference for IBM professionals, IBM Business Partners, and technical specialists who support the p5-590 and p5-595, and for those who want to learn more about this radically new server in a clear, single-source handbook.

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.

Ding Wei is an Advisory IT Specialist working for IBM China ATS. He has eight years of experience in the Information Technology field. His areas of expertise include pSeries and storage products and solutions. He has been working for IBM for six years.

Peter Domberg (Domi) is a Technical Support Specialist in Germany. He has 27 years of experience in the ITS hardware service. His areas of expertise include pSeries, RS/6000®, networking, and SSA storage. He is also an AIX certified specialist and hardware support specialist for the North and East Germany regions.

TaiJung Kim is a pSeries systems product engineer at the pSeries post-sales technical support team in IBM Korea. He has three years of experience working on RS/6000 and pSeries products. He is an IBM Certified Specialist in pSeries systems and AIX. He provides clients with technical support on pSeries systems, AIX, and system management.

Nia Kelley is a Staff Software Engineer based in IBM Austin with over four years of experience in the pSeries firmware development field. She holds a Bachelors of Science degree in Electrical Engineering from the University of Maryland at College Park. Her areas of expertise include system bringup and firmware development. She has led several project teams working in her areas of expertise, in addition to holding various architectural positions for existing and future pSeries products. Ms. Kelley is an alumni of the IBM Extreme Blue program and has filed numerous patents for the IBM corporation.

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

International Technical Support Organization, Austin Center
Scott Vetter

IBM Austin
Anis Abdul, George Ahrens, Doug Bossen, Pat Buckland, Mark Dewalt,
Bob Foster, Iggy Haider, Dan Henderson, Richard (Jamie) Knight,
Andy McLaughlin, Cathy Nunez, Jayesh Patel, Craig Shempert, Guillermo Silva,
Joel Tendler

IBM Endicott
Brian Tolan

IBM Raleigh
Andre Metelo

IBM Rochester
Salim Agha, Diane Knipfer, Dave Lewis, Matthew Spinler, Stephanie Swanson

IBM Poughkeepsie
Doug Baska

IBM Boca Raton
Arthur J. Prchlik

IBM Somers
Bill Mihaltse

IBM UK
Derrick Daines, Dave Williams

IBM France
Jacques Noury

IBM Germany
Hans Mozes, Wolfgang Seiwald

IBM Australia
Cameron Ferstat

BM Italy
Carlo Costantini

IBM Redbook “Partitioning Implementations for IBM eserver p5 and pSeries Servers” Team
Nic Irving (CSC Corporation - Australia), Matthew Jenner (IBM Australia),
Arsi Kortesnemi (IBM Finland)

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 customers.

Your efforts will help increase product acceptance and customer 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

System overview

In this chapter we provide a basic overview of the p5-590 and p5-595 servers, highlighting the new features, marketing position, main features, and operating systems.

- Section 1.1, “Introduction” on page 2
- Section 1.2, “What’s new” on page 2
- Section 1.3, “General overview and characteristics” on page 4
- Section 1.4, “Features summary” on page 12
- Section 1.5, “Operating systems support” on page 13

1.1 Introduction

The IBM @server p5 590 and IBM @server p5 595 are the servers 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 and twice the capacity of its predecessor, the IBM @server 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.

Both systems are powered by IBMs most advanced 64-bit Power Architecture microprocessor, the IBM POWER5 microprocessor, with simultaneous multi-threading that makes each processor function as two to the operating system, thus increasing commercial performance and system utilization over servers without this capability. The p5-595 features a choice of IBMs fastest POWER5 processors running at 1.90 GHz or 1.65 GHz, while the p5-590 offers 1.65 GHz processors.

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

1.2 What’s new

The p5-590 and p5-595 bring the following features:

- **POWER5 microprocessor**

  Designed to provide excellent application performance and high reliability. Includes simultaneous multi-threading to help increase commercial system performance and processor utilization. See Section 1.3.1, “Microprocessor technology” on page 6 and Section 2.2, “The POWER5 microprocessor” on page 18 for more information.

- **High memory / I/O bandwidth**

  Fast processors wait less for data to be moved through the system. Delivers data faster for the needs of high performance computing and other memory-intensive applications. See Section 2.3, “Memory subsystem” on page 26 for more information.

* Indicates this feature is optional, is available on selected models, or requires separate software.

- **Flexibility in packaging**

  High-density 24-inch system frame for maximum growth. See Section 1.3, “General overview and characteristics” on page 4 for more information.

- **Shared processor pool***

  Provides the ability to transparently share processing power between partitions. Helps balance processing power and ensures the high priority partitions receive the processor cycles they need. See Section 3.2.1, “Shared processor partitions” on page 58 for more information.

- **Micro-Partitioning***

  Allows each processor in the shared processor pool to be split into as many as ten partitions. Fine-tuned processing power to match workloads. See Section 3.2, “Micro-Partitioning” on page 57 for more information.

- **Virtual I/O***

  Shares expensive resources to help reduce costs. See Section 3.6, “Virtual SCSI” on page 80 for more information.

- **Virtual LAN***

  Provides the capability for TCP/IP communication between partitions without the need for additional network adapters. See Section 3.3, “Virtual Ethernet” on page 64 for more information.

- **Dynamic logical partitioning**

  Allows reallocation of system resources without rebooting affected partitions. Offers greater flexibility in using available capacity and more rapidly matching resources to changing business requirements.

- **Mainframe-inspired RAS**

  Delivers exceptional system availability using features usually found on much more expensive systems including service processor, Chipkill memory, first failure data capture, dynamic deallocation of selected system resources, dual system clocks, and more. See Chapter 6, “Reliability, availability, and serviceability” on page 137 for more information.

- **Broad range of CoD offerings***

  Provides temporary access to processors and memory to meet predictable business spikes. Prepaid access to processors to meet intermittent or seasonal demands. Offers a one-time 30 day trial to test increased processor or memory capacity before permanent activation. Allow processors and memory to be permanently added to meet long term workload increases. See Chapter 4, “Capacity on Demand” on page 83 for more information.

- **Grid Computing support***

  Allows sharing of a wide range of computing and data resources across heterogeneous, geographically dispersed environments.

- **Scaling through Cluster Systems Management support**

  Allows for more granular growth so end-user demands can be readily satisfied. Provides centralized management of multiple interconnected systems. Provides ability to handle unexpected workload peaks by sharing resources.

- **Multiple operating system support**

  Allows clients the flexibility to select the right operating system and the right application to meet their needs. Provides the ability to expand applications choices to include many open source applications. See Section 1.5, “Operating systems support” on page 13for more information.

1.3 General overview and characteristics

The p5-590 and p5-595 servers are designed with a basic server configuration that starts with a single rack, sometimes called a *frame* (Figure 1-1), and is featured with optional and required components.

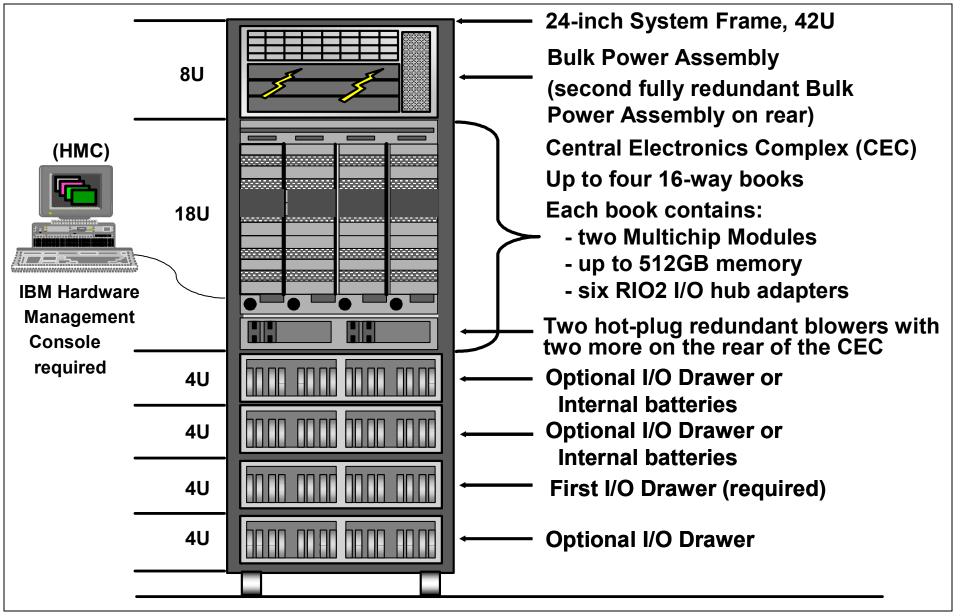
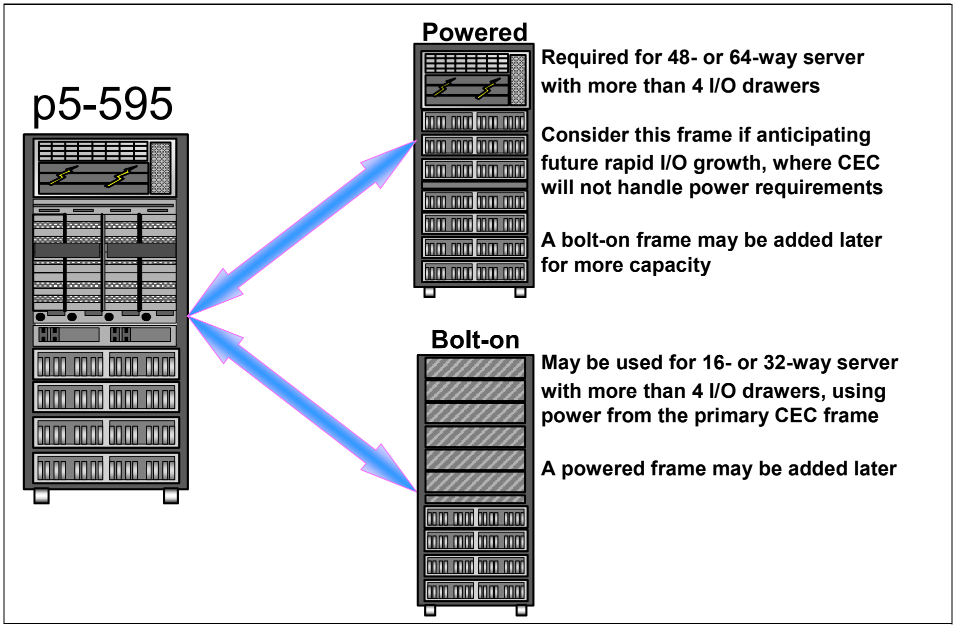

24-inch System Frame, 42U
8U
Bulk Power Assembly
(second fully redundant Bulk Power Assembly on rear)
(HMC)
18U
Central Electronics Complex (CEC)
Up to four 16-way books
Each book contains:
- two Multichip Modules
- up to 512GB memory
- six RIO2 I/O hub adapters
IBM Hardware Management Console required
Two hot-plug redundant blowers with two more on the rear of the CEC
4U
Optional I/O Drawer or Internal batteries
4U
Optional I/O Drawer or Internal batteries
4U
First I/O Drawer (required)
4U
Optional I/O Drawer

Figure 1-1 Primary system frame organization

Both systems are powered by IBMs most advanced 64-bit Power Architecture microprocessor, the IBM POWER5 microprocessor, with simultaneous multi-threading that makes each processor logically appear as two to the operating system, thus increasing commercial throughput and system utilization over servers without this capability. The p5-595 features a choice of IBMs fastest POWER5 microprocessors running at 1.9 GHz or 1.65 GHz, while the p5-590 offers 1.65 GHz processors.

For additional capacity, either a powered or non-powered frame can be configured for a p5-595, as shown in .



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 1-2 Powered and bolt on frames

The p5-590 can be expanded by an optional bolt-on frame.

Every p5-590 and p5-595 server comes standard with Advanced POWER Virtualization, providing Micro-Partitioning, Virtual I/O Server, and Partition Load Manager (PLM). Micro-Partitioning enables system configurations with more partitions than processors. Processing resources can be allocated in units as small as 1/10th of a processor and be fine-tuned in increments of 1/100th of a processor. So a p5-590 or p5-595 system can define up to ten *virtual servers* per processor (maximum of 254 per system), controlled in a shared processor pool for automatic, non-disruptive resource balancing. Virtualization features of the p5-590 and the p5-595 are introduced in Chapter 3, “POWER5 virtualization capabilities” on page 55.

The ability to communicate between partitions using virtual Ethernet is part of the Advanced POWER Virtualization feature and it is extended with the Virtual I/O Server to include shared Ethernet adapters. Also part of the Virtual I/O Server is virtual SCSI for sharing SCSI adapters and the attached disks.

The Virtual I/O Server requires APAR IY62262 and is supported by AIX 5L Version 5.3 with APAR IY60349, as well as by SLES 9 and RHEL AS 3. Also included in Advanced POWER Virtualization is PLM, a powerful policy based tool for automatically managing resources among LPARs running AIX 5L Version 5.3 or AIX 5L Version 5.2 with the 5200-04 Recommended Maintenance package.

IBM @server p5 590 and p5 595 servers also offer optional Capacity on Demand (CoD) capability for processors and memory. CoD functionality is outlined in Chapter 4, “Capacity on Demand” on page 83.

IBM @server p5 590 and p5 595 servers provide significant extensions to the mainframe-inspired reliability, availability, and serviceability (RAS) capabilities found in IBM @server p5 and pSeries systems. They come equipped with multiple resources to identify and help resolve system problems rapidly. During ongoing operation, error checking and correction (ECC) checks data for errors and can correct them in real time. First Failure Data Capture (FFDC) capabilities log both the source and root cause of problems to help prevent the recurrence of intermittent failures that diagnostics cannot reproduce. Meanwhile, Dynamic Processor Deallocation and dynamic deallocation of PCI bus slots help to reallocate resources when an impending failure is detected so applications can continue to run unimpeded. RAS function is discussed in Chapter 6, “Reliability, availability, and serviceability” on page 137.

Power options for these systems are described in Section 5.2.10, “Rack, power, and battery backup configuration rules” on page 128.

A description of RAS features, such redundant power and cooling, can be found in Section 6.4, “Redundancy in components” on page 144.

The following sections detail some of the technologies behind the p5-590 and p5-595.

1.3.1 Microprocessor technology

The IBM POWER4 microprocessor, which was introduced in 2001, was a result of advanced research technologies developed by IBM to create a high-performance, high-scalability chip design to power future IBM @server systems. The POWER4 design integrates two processor cores on a single chip, a shared second-level cache, a directory for an off-chip third-level cache, and the necessary circuitry to connect it to other POWER4 chips to form a system. The dual-processor chip provides natural thread-level parallelism at the chip level.

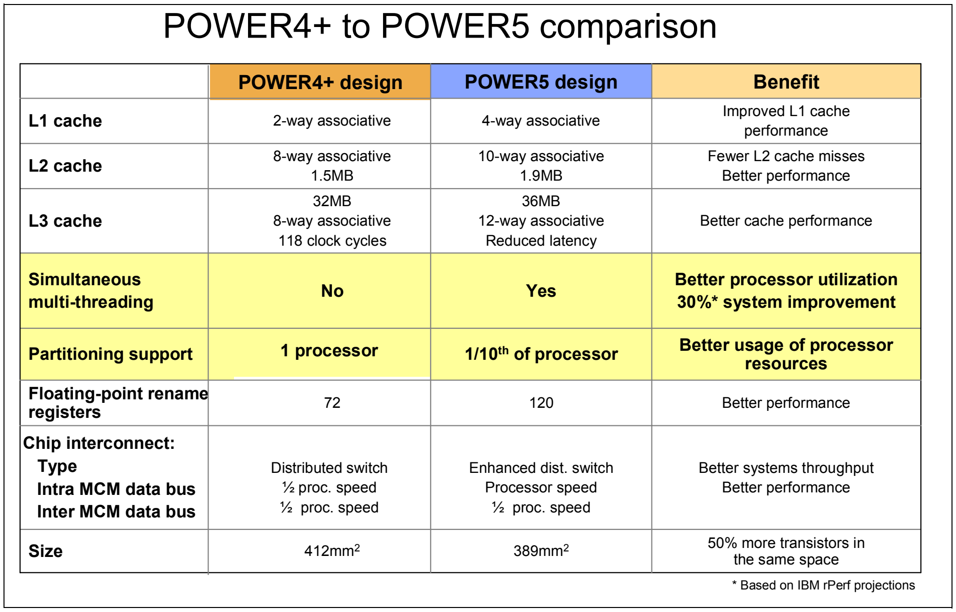
The POWER5 microprocessor is IBMs second generation dual core microprocessor and extends the POWER4 design by introducing enhanced performance and support for a more granular approach to computing. The POWER5 chip features single- and multi- threaded execution and higher performance in the single-threaded mode than the POWER4 chip at equivalent frequencies.

The primary design objectives of the POWER5 microprocessor are:

- Maintain binary and structural compatibility with existing POWER4 systems
- Enhance and extend symmetric multiprocessing (SMP) scalability
- Continue to provide superior performance
- Deliver a power efficient design
- Enhance reliability, availability, and serviceability

POWER4 to POWER5 comparison

There are several major differences between POWER4 and POWER5 chip designs, and they include the following areas shown in Figure 1-3, and as discussed in the following sections:



POWER4+ to POWER5 comparison

	POWER4+ design	**POWER5 design**	**Benefit**
L1 cache	2-way associative	4-way associative	Improved L1 cache performance
L2 cache	8-way associative 1.5MB	10-way associative 1.9MB	Fewer L2 cache misses Better performance
L3 cache	32MB 8-way associative 118 clock cycles	36MB 12-way associative Reduced latency	Better cache performance
Simultaneous multi-threading	**No**	**Yes**	**Better processor utilization** **30%* system improvement**
Partitioning support	**1 processor**	**1/10th of processor**	**Better usage of processor resources**
Floating-point rename registers	72	120	Better performance
Chip interconnect: **Type** **Intra MCM data bus** **Inter MCM data bus**	 Distributed switch ½ proc. speed ½ proc. speed	 Enhanced dist. switch Processor speed ½ proc. speed	Better systems throughput Better performance
Size	412mm^2	389mm^2	50% more transistors in the same space

* Based on IBM rPerf projections

Figure 1-3 POWER4 and POWER5 architecture comparison

Introduction to simultaneous multi-threading

Simultaneous multi-threading is a hardware design enhancement in POWER5 Architecture that allows two separate instruction streams (threads) to execute simultaneously on the processor. It combines the capabilities of super scaler processors with the latency hiding abilities of hardware multithreading.

Using multiple on-chip thread contexts, the simultaneous multi-threading processor executes instructions from multiple threads each cycle. By duplicating portions of logic in the instruction pipeline and increasing the capacity of the register rename pool, the POWER5 processor can execute several elements of two instruction streams, or threads, concurrently. Through hardware and software thread prioritization, greater utilization of the hardware resources can be realized without an impact to application performance.

The benefit of simultaneous multi-threading is realized more in commercial environments over numeric intensive environments, since the number of transactions performed outweighs the actual speed of the transaction. For example, the simultaneous multi-threading environment would be much better suited for a Web server or database server than it would be for a Fortran weather prediction application. In the rare case that applications are tuned to optimize the use of processor resources and see a decrease in performance due to increased contention to cache and memory, simultaneous multi-threading may be disabled.

Although it is the operating system that determines whether simultaneous multi-threading is used, simultaneous multi-threading is otherwise completely transparent to the applications and operating system, and implemented entirely in hardware.

1.3.2 Memory subsystem

With the enhanced architecture of larger 7.6 MB L2 and 144 MB L3 caches, each mutichip module (MCM) can stage information more effectively from processor memory to applications. These caches allow the p5-590 and p5-595 to run workloads significantly faster than predecessor servers.

The difference of memory hierarchy between POWER4 and POWER5 systems is represented in Figure 1-4 as follows:

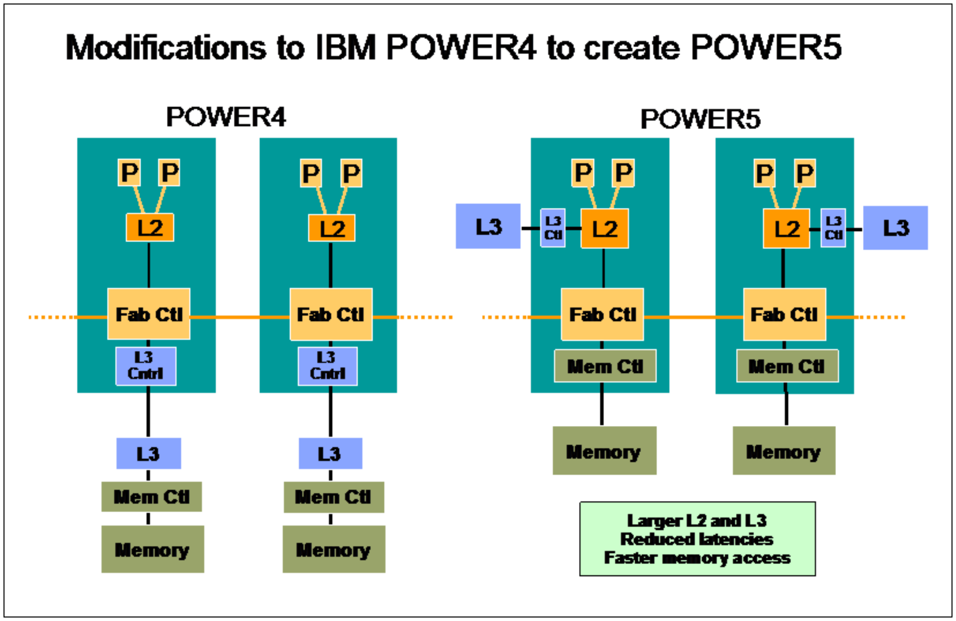


Figure 1-4 POWER4 and POWER5 memory structure comparison

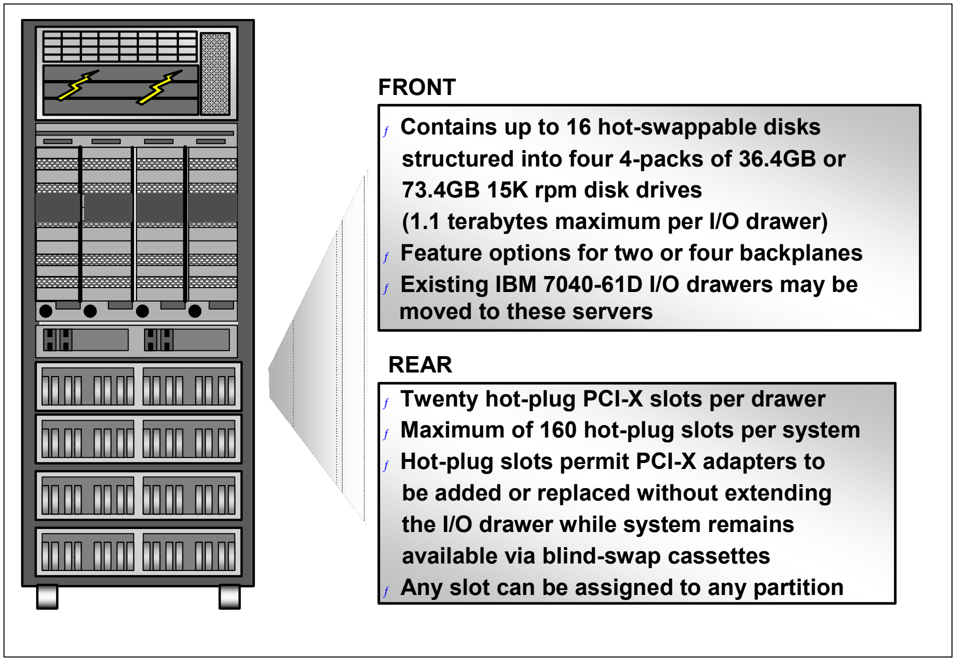
There are two types of memory technologies offered, namely DDR1 and DDR2. Equipped with 8 GB of memory in its minimum configuration, the p5-590 can be scaled to 1 TB using DDR1 266 MHz memory. From 8 GB to 128 GB of DDR2 533 MHz memory, useful for high-performance applications, is available. The p5-595 can be scaled from 8 GB to 2 TB of DDR1 266 MHz memory; From 8 GB to 256 GB of DDR2 533 MHz memory (at the time of writing).

Additional information about memory can be found in Chapter 2.3, “Memory subsystem” on page 26 and Section 5.2.4, “Memory configuration rules” on page 121.

1.3.3 I/O subsystem

Using the RIO-2 ports in the processor books, up to twelve I/O drawers can be attached to a p5-595 and up to eight I/O drawers to the p5-590, providing up to 9.3 TB and 14 TB of 15 K RPM disk storage, respectively. Each 4U (4 EIA unit) drawer provides 20 hot plug, blind-swap PCI-X I/O adapter slots, 16 front-accessible, hot-swappable disk drive bays and four integrated Ultra3 SCSI controllers. I/O drawers can be installed in the primary 24-inch rack or in an optional expansion rack. Attachment to a wide range of IBM TotalStorage storage system offerings – including disk storage subsystems, storage area network components, tape libraries, and external media drives – is also supported.

A minimum of one I/O drawer (FC 5791 or FC 5794) is required per system. I/O drawer FC 5791 contains 20 PCI-X slots and 16 disk bays, and FC 5794 contains 20 PCI-X slots and 8 disk bays. Existing 7040-61D I/O drawers may also be attached to a p5-595 or p5-590 servers as additional I/O drawers (when correctly featured). For more information on the I/O system, refer to Chapter 2.7, “I/O drawer” on page 35. The I/O features are shown in Figure 1-5.



FRONT

- **Contains up to 16 hot-swappable disks structured into four 4-packs of 36.4GB or 73.4GB 15K rpm disk drives (1.1 terabytes maximum per I/O drawer)**
- **Feature options for two or four backplanes**
- **Existing IBM 7040-61D I/O drawers may be moved to these servers**

REAR

- **Twenty hot-plug PCI-X slots per drawer**
- **Maximum of 160 hot-plug slots per system**
- **Hot-plug slots permit PCI-X adapters to be added or replaced without extending the I/O drawer while system remains available via blind-swap cassettes**
- **Any slot can be assigned to any partition**

Figure 1-5 p5-590 and p5-595 I/O drawer organization

1.3.4 Virtualization

The IBM Virtualization Engine can help simplify IT infrastructure by reducing management complexity and providing integrated virtualization technologies and systems services for a single IBM eserver p5 server or across multiple server platforms. It brings together existing offerings and new technologies. The IBM Virtualization Engine systems technologies added or enhanced the systems using the POWER5 Architecture are as follows:

- **POWER Hypervisor**

  Is responsible for time slicing and dispatching the logical partition workload across the physical processors. The POWER Hypervisor also enforces partition security, and can provide Virtual LAN channels between partitions, reducing the need for physical Ethernet adapters using I/O adapter slots.

- **Simultaneous multi-threading**

  Allows two separate instruction streams (threads) to run concurrently on the same physical processor, improving overall throughput and improving overall hardware resource utilization.

- **Dynamic LPAR**

  Allows system resources (processors, memory, and I/O adapters and attached devices) to be grouped logically into separate systems within the same server.

- **Micro-Partitioning**

  Allows processor resources to be allocated to partitions in units as small as 1/10th of a processor, with increments in units of 1/100th of a processor.

- **Virtual I/O**

  Includes Virtual SCSI for sharing SCSI attached disks and virtual networking to enable sharing of Ethernet adapters.

- **Virtual LAN (VLAN)**

  Enables high-speed, secure, partition-to-partition communications using the TCP/IP protocol to help improve performance.

- **Capacity on Demand**

  Allows system resources such as processors and memory to be made available on an as-needed basis.

- **Multiple operating system support**

  The POWER5 processor-based @server p5 products supports IBM AIX 5L Version 5.2, IBM AIX 5L Version 5.3, SUSE Linux Enterprise Server 9 (SLES9), and Red Hat Enterprise AS Linux 3 (RHEL AS 3). IBM i5/OS V5R3 is also available on @server p5 models 570, 590, and 595.

A detailed description of these features can be found in Chapter 3, “POWER5 virtualization capabilities” on page 55.

1.4 Features summary

Table 1-1 summarizes the major features of the p5-590 and p5-595 servers. For mroe information, see Appendix A, “Facts and features reference” on page 245.

Table 1-1 p5-590 and p5-595 features summary

IBM @server p5 system	p5-590	p5-595
Machine type - Model	9119-590	9119-595
Packaging	24-inch system frame	
Number of expansion frames	1	1 or 2
Number of processors per system	8 to 32	16 to 64
POWER5 processor speed	1.65 GHz	1.65 or 1.90 GHz
Number of 16-way processor books (2 MCMs per book)	1 or 2	1, 2, 3, or 4
Memory	8 GB - 1024 GB*	8 GB - 2048 GB*
CoD features (All memory CoD features apply to DDR1 memory only)	Processor/Memory CUoD Reserve CoD On/Off Processor/Memory CoD Trial Processor/Memory CoD Capacity BackUp	
Maximum micro-partitions	10 times the number of processors (254 maximum)	
PCI-X slots	20 per I/O drawer	
Media bays	Optional	
Disk bays	16 per I/O drawer	
Optional I/O drawers	Up to 8	Up to 12
Maximum PCI-X slots with maximum I/O drawers	160	240
Maximum disk bays with maximum I/O drawers	128	192
Maximum disk storage maximum with I/O drawers	9.3 TB	14.0 TB

* 32 GB memory cards to enable maximum memory are planned for availability April 8, 2005. Until that time, maximum memory is half as much (512 GB on p5-590 and 1024 GB on p5-595).

1.5 Operating systems support

All new POWER5 processor-based servers are capable of running IBM AIX 5L Version 5.3 or AIX 5L Version 5.2 for POWER and support appropriate versions of Linux. Both of the aforementioned supported versions of AIX 5L have been specifically developed and enhanced to exploit and support the extensive RAS features on IBM @server pSeries systems. Table 1-2 lists operating systems compatibility.

Table 1-2 p5-590 and p5-595 operating systems compatibility

Operating system	p5-590	p5-595
AIX 5L V5.1	No	No
AIX 5L V5.2(5765-E62)	Yes	Yes
AIX 5L V5.3(5765-G03)	Yes	Yes
AIX 5L LPAR	Yes	Yes
Red Hat Enterprise Linux AS 3 for POWER (5639-RDH)	Yes	Yes
SUSE LINUX Enterprise Server 8	No	No
SUSE LINUX Enterprise Server 9 for POWER (5639-SLP)	Yes	Yes
Linux LPAR	Yes	Yes
i5/OS	Yes	Yes
HACMP for AIX 5L V5.2 (5765-F62)	Yes	Yes
Cluster Systems Management for AIX 5L V1.4 (5765-F67)	Yes	Yes
Cluster Systems Management for Linux on POWER V1.4 (5765-G16)	Yes	Yes

[1] Many of the features described in this document are operating system dependent and may not be available on Linux. For more information, check: http://www.ibm.com/servers/eserver/pseries/linux/whitepapers/linux_pseries.html

1.5.1 AIX 5L

The p5-590 and p5-595 requires AIX 5L Version 5.3 or AIX 5L Version 5.2 Maintenance Package 5200-04 (IY56722) or later.

The system requires the following media:

- AIX 5L for POWER V5.2 5765-E62, dated 08/2004, or later (CD# LCD4-1133-04) plus APAR IY60347 (Required AIX 5.2 updates for @server p5 590/595)
- AIX 5L for POWER Version 5.3 5765-G03, dated 08/2004, or later. (CD# LCD4-7463-00) with APAR IY60349 (Required AIX 5.3 updates for @server p5 590/595)

IBM periodically releases maintenance packages for the AIX 5L operating system. These packages are available on CD-ROM (FC 0907) and can be downloaded from the Internet at:

http://www.ibm.com/servers/eserver/support/pseries/aixfixes.html

You can also get individual operating system fixes and information about obtaining AIX 5L service at this site. AIX 5L Version 5.3 has the Service Update Management Assistant (SUMA) tool, which helps the administrator to automate the task of checking and downloading operating system downloads.

The Advanced POWER Virtualization feature is not supported on AIX 5L Version 5.2. AIX 5L Version 5.3 is required to take full advantage of Advanced POWER Virtualization feature.

1.5.2 Linux

For the p5-590 and p5-595, Linux distributions are available through SUSE and Red Hat at the time this publication was written. The p5-590 and p5-595 requires the following version of Linux distributions:

- SUSE LINUX Enterprise Server 9 for POWER, or later
- Red Hat Enterprise Linux AS 3 for POWER, or later

The Advanced POWER Virtualization feature, DLPAR, and other features require SUSE SLES 9.

In Japan, Turbolinux is also available. In the Latin America sales region, Conectiva is also available. For related information and an overview, see:

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

Find full information about Red Hat Enterprise Linux AS 3 for POWER at:

http://www.redhat.com/software/rhel/as/

Find full information about SUSE Linux Enterprise Server 9 for POWER at:

http://www.suse.com/us/business/products/server/sles/i_pseries.html

For information about UnitedLinux for pSeries from Turbolinux, see:

http://www.turbolinux.co.jp

For the latest in IBM Linux news, subscribe to the Linux Line. See:

https://www6.software.ibm.com/reg/linux/linuxline-i

Many of the features that are described in this document are OS-dependent and may not be available on Linux. For more information, check:

http://www.ibm.com/servers/eserver/pseries/linux/whitepapers/linux_pseries.html

IBM only supports the Linux systems of customers with a SupportLine contract that covers Linux. Otherwise, the Linux distributor should be contacted for support.

2

Hardware architecture

This chapter reviews the contents in Chapter 1, “System overview” on page 1, but provides deeper technical descriptions of the topics, including hardware architectures that are implemented in the p5-590 and p5-595 servers, the POWER5 processor, memory subsystem, and I/O subsystem in the following topics:

- Section 2.1, “Server overview” on page 18
- Section 2.2, “The POWER5 microprocessor” on page 18
- Section 2.3, “Memory subsystem” on page 26
- Section 2.4, “Central electronics complex” on page 29
- Section 2.5, “System flash memory configuration” on page 34
- Section 2.6, “Vital product data and system smart chips” on page 35
- Section 2.7, “I/O drawer” on page 35

2.1 Server overview

The IBM @server p5 595 and p5 590 provides an expandable, high-end enterprise solution for managing the computing requirements needed to become an on demand business. With the introduction of the POWER5 Architecture, there has been numerous improvements over the previous POWER4 architecture based systems.

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

The p5-595 is a 16/32/48/64-way (at 1.9 GHz or 1.65 GHz) SMP system packaged in a 24-inch wide 18 EIA by 36 inch deep CEC. The CEC is installed in the 42 EIA base primary rack that also include two top mounted front and back bulk power assemblies (BPAs) and support for up to four I/O drawers. A powered I/O rack (FC 5792) and a bolt-on expansion frame (FC 8691) is also available to support additional I/O drawers for the p5-595 system. Up to 12 I/O drawers can be attached to a p5-595.

The p5-590 has identical architecture with p5-595. It differs from p5-595 in the following areas:

- Only 1.65 GHz processor are support in a p5-590.
- The maximum configuration is a 32-way system with up to eight I/O drawers.
- A powered I/O rack (FC 5792) is not required in the p5-590.

2.2 The POWER5 microprocessor

The POWER5 processor features single-threaded and multi-threaded execution, providing higher performance in the single-threaded mode than its POWER4 predecessor provides at equivalent frequencies. The POWER5 microprocessor maintains both binary and architectural compatibility with existing POWER4 systems to ensure that binaries continue executing properly and that all application optimizations carry forward to newer systems. The POWER5 microprocessor provides additional enhancements such as virtualization, simultaneous multi-threading support, improved reliability, availability, and serviceability at both chip and system levels, and it has been designed to support interconnection of 64 processors along with higher clock speeds.

Figure 2-1 shows the high-level structures of POWER4 and POWER5 processor-based systems. The POWER4 processors scales up to a 32-way symmetric multi-processor. Going beyond 32 processors with POWER4 Architecture could increase interprocessor communication, resulting in higher traffic on the interconnection fabric bus. This can cause greater contention and negatively affect system scalability.

Moving the L3 cache reduces traffic on the fabric bus and enables POWER5 processor-based systems to scale to higher levels of symmetric multi-processing. The POWER5 processor supports a 1.9 MB on-chip L2 cache, implemented as three identical slices with separate controllers for each. Either processor core can independently access each L2 controller. The L3 cache, with a capacity of 36 MB, operates as a backdoor with separate buses for reads and writes that operate at half processor speed.

Because of the higher transistor density of the POWER5 0.13-µm technology, it was possible to move the memory controller on-chip and eliminate a chip that was previously needed for the memory controller function. These changes in the POWER5 processor also have the significant side benefits of reducing latency to the L3 cache and main memory, as well as reducing the number of chips that are necessary to build a system.

The POWER5 processor supports the 64-bit PowerPC architecture. A single die contains two identical processor cores, each supporting two logical threads. This architecture makes the chip appear as a four-way symmetric multi-processor to the operating system. The POWER5 processor core has been designed to support both enhanced simultaneous multi-threading and single-threaded (ST) operation modes.

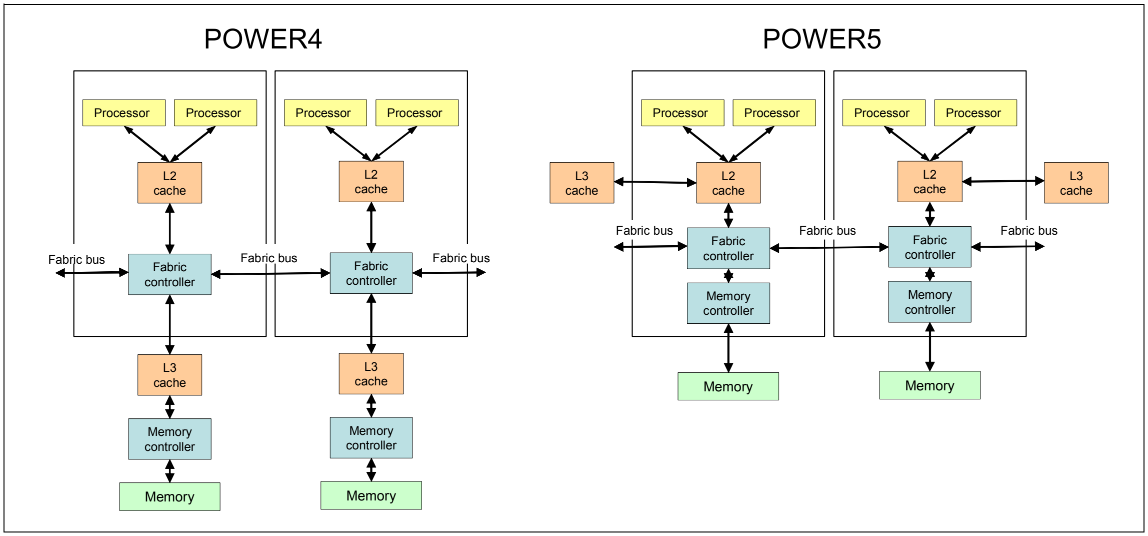


Figure 2-1 POWER4 and POWER5 system structures

2.2.1 Simultaneous multi-threading

As a permanent requirement for performance improvements at the application level, simultaneous multi-threading functionality is embedded in the POWER5 chip technology. Developers are familiar with process-level parallelism (multi-tasking) and thread-level parallelism (multi-threads). simultaneous multi-threading is the next stage of processor for achieving higher processor utilization for throughput-oriented applications to introduce the method of instruction group-level parallelism to support multiple pipelines to the processor. The instruction groups are chosen from different hardware threads belonging to a single OS image.

simultaneous multi-threading is activated by default when an OS that supports it is loaded. On a 2-way POWER5 processor-based system, the operating system discovers the available processors as a 4-way system. To achieve a higher performance level, simultaneous multi-threading is also applicable in Micro-Partitioning, capped or uncapped, and dedicated partition environments.

Simultaneous multi-threading is supported on POWER5 processor-based systems running AIX 5L Version 5.3 or Linux-based systems at a required 2.6 kernel. AIX provides the **smtctl** command that turns simultaneous multi-threading on and off without subsequent reboot. For Linux, an additional boot option must be set to activate simultaneous multi-threading after a reboot.

The simultaneous multi-threading mode increases the usage of the execution units. In the POWER5 chip, more rename registers have been introduced (both Floating Point registers (FPR) and general-purpose registers (GPR) are increased to 120), that are essential for out-of-order execution and vital for the simultaneous multi-threading.

Enhanced simultaneous multi-threading features

To improve simultaneous multi-threading performance for various workload mixes and provide robust quality of service, POWER5 provides two features:

- Dynamic resource balancing

  The objective of dynamic resource balancing is 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 a different thread throttling mechanism.

- Adjustable thread priority

  Adjustable thread priority lets software 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.

Single threaded operation

Not all applications benefit from simultaneous multi-threading. Having threads executing on the same processor does not increase the performance of processor intensive applications or applications that consume all of the chip's memory bandwidth. For this reason, the POWER5 processor supports the single thread (ST) execution mode. In this mode, the POWER5 processor gives all of the physical resources to the active thread, enabling it to achieve higher performance than a POWER4 processor-based system at equivalent frequencies. Highly optimized scientific codes are one example where ST operation is ideal.

simultaneous multi-threading and ST operation modes can be dynamically switched without affecting server operations. The two modes can coexist on a single physical system; however, only a single mode is possible on each OS instance (partition).

2.2.2 Dynamic power management

In current CMOS[1] technologies, chip power consumption is one of the most important design parameters. With the introduction of simultaneous multi-threading, more instructions execute per cycle per processor core, thus increasing the core's and the chip's total switching power. To reduce switching power, POWER5 chips extensively use a fine-grained, dynamic clock-gating mechanism. This mechanism gates off clocks to a local clock buffer if dynamic power management logic knows that the set of latches that are driven by the buffer will not be used in the next cycle. This allows substantial power saving with no performance impact. In every cycle, the dynamic power management logic determines whether a local clock buffer that drives a set of latches can be clock-gated in the next cycle.

In addition to the switching power, leakage power has become a performance limiter. To reduce leakage power, the POWER5 chip uses transistors with low threshold voltage only in critical paths. The POWER5 chip also has a low-power mode, enabled when the system software instructs the hardware to execute both threads at priority 1. In low power mode, instructions dispatch once every 32 cycles at most, further reducing switching power. Both threads are set to priority 1 by the operating system when in the idle loop.

2.2.3 The POWER chip evolution

The p5-590 and p5-595 system complies with the RS/6000 platform architecture, which is an evolution of the PowerPC Common Hardware Reference Platform (CHRP) specifications. Figure 2-2 on page 23 shows the POWER chip evolution.

[1] Complementary Metal Oxide Semiconductor

- **POWER4**

  POWER4 processor is not just a chip, but rather an architecture of how a set of chips is designed together to build a system. As such, POWER4 can be considered a technology in its own right. The interconnect topology, referred to as a Distributed Switch, was new to the industry with POWER4. In that light, systems are built by interconnecting POWER4 chips to form up to 32-way symmetric multi-processors. The reliability, availability, and serviceability (RAS) design incorporated into POWER4 is pervasive throughout the system and is as much a part of the design. POWER4 is the chip technology used in the pSeries Models 615, 630, 650, 655, 670, 690, and Intellistation 275. It is also the basis for the Power PC 970 used in JS20 BladeCenter servers.

  The POWER4 design can handle a varied and robust set of workloads. This is especially important as the on demand business world evolves and data intensive demands on systems merge with commercial requirements. The need to satisfy high performance computing requirements with its historical high bandwidth demands and commercial requirements, along with data sharing and SMP scaling requirements dictate a single design to address both environments.

- **POWER5**

  POWER5 technology is the next generation of 64-bit architecture. Although the hardware is based on POWER4, POWER5 is much more than just an improvement in processor or chip design. It is a major architectural change, creating a much more efficient superscalar processor complex. For example, the high performance distributed switch is enhanced. POWER5 technology is implemented in the @server p5 Models 510, 520, 550, 570, 575, 590, 595 and the OpenPower 710 and 720.

  As with POWER4 hardware technology, POWER5 technology-based processors have two load/store, two arithmetic, one branch execution unit, and one execution unit for logical operations on the cycle redundancy (CR). The design of the processor complex is such that it can most efficiently execute multiple instruction streams concurrently. With simultaneous multi-threading active, instructions from two different threads can be issued per single cycle.

  The POWER5 concept is a step further into autonomic computing[2]. Several enhanced reliability and availability enhancements are implemented. Along with increased redundant components, it incorporates new technological high standards, such as special ways to reduce junction temperatures to reach a high level of availability. The full system design approach is required to

[2] Autonomic computing: An approach to self-managed computing systems with a minimum of human interference. The term derives from the body's autonomic nervous system, which controls key functions without conscious awareness or involvement.

maintain balanced utilization of hardware resources and high availability of the new @server p5 systems.

Memory and CPU sharing, a dual clock, and dual service processors with failover capability are examples of the full system design approach for high availability. IBM designed the @server p5 system processor, caching mechanisms, memory allocation methods, and full ECC support for buses between chips inside a POWER5 system for performance and availability. In addition, advanced error correction and low power consumption circuitry is improved with thermal management.

Multi-processor POWER5 technology-based servers have multiple autonomic computing features for higher availability compared with single processor servers. If a processor is running, but is experiencing a high rate of correctable soft errors, it can be deconfigured. Its workload can be picked up automatically by the remaining processor or processors without an IPL. If there is an unused Capacity Upgrade on Demand processor or if one processor unit of unused capacity in a shared processor pool is available, the deconfigured processor can be replaced dynamically by the unused processor capacity to maintain the same level of available performance.

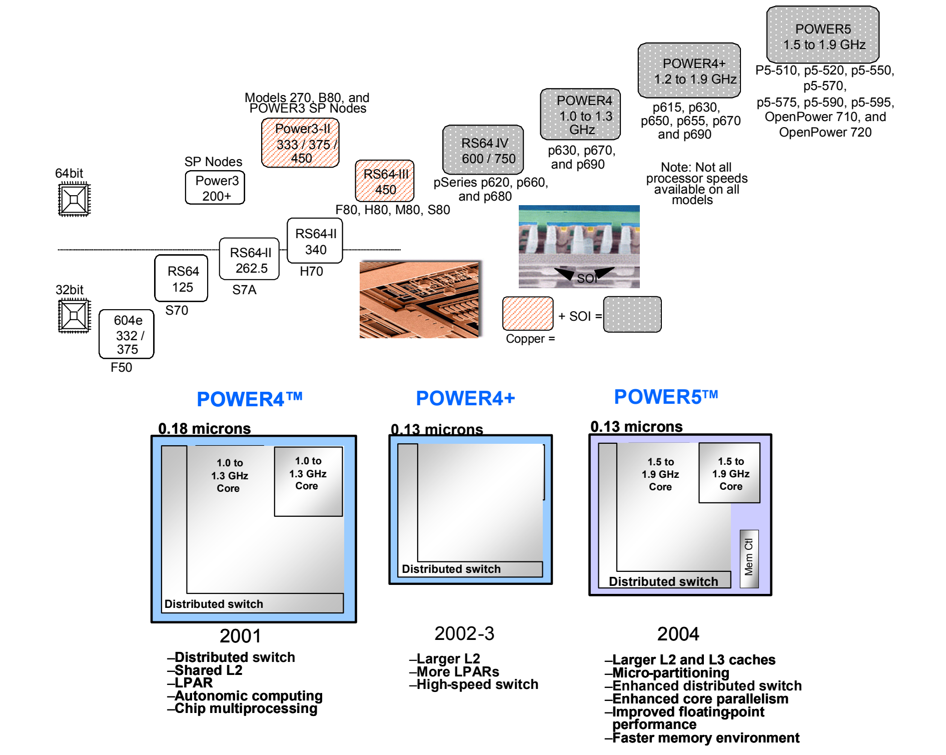


Figure 2-2 The POWER chip evolution

2.2.4 CMOS, copper, and SOI technology

The POWER5 processor design enables IBM eserver p5 systems to offer clients improved performance, reduced power consumption, and decreased IT footprint size through logical partitioning, Micro-Partitioning and Virtual I/O. The POWER5 processor chip takes advantage of IBM leadership technology. It is made using IBM 0.13-µm-lithography CMOS. The POWER5 processor also uses copper and Silicon-on-Insulator (SOI) technology to allow a higher operating frequency for improved performance, yet with reduced power consumption and improved reliability compared to processors not using this technology.

2.2.5 Processor books

In the p5-590 and p5-595 system, the POWER5 chip has been packaged with the L3 cache chip into a cost-effective multi chip module (MCM) package. The storage structure for the POWER5 processor chip is a distributed memory architecture that provides high-memory bandwidth. Each processor can address all memory and sees a single shared memory resource. As such, two MCMs with their associated L3 cache and memory are packaged on a single processor book. Access to memory behind another processor is accomplished through the fabric buses. The p5-590 supports up to two processor books (each book is a 16-way) and the p5-595 supports up to four processor books. Each processor book has dual MCMs containing POWER5 processor chips and 36 MB L3 modules. Each 16-way processor book also includes 16 slots for memory cards and six remote I/O-2 (RIO-2) attachment cards for connection of the system I/O drawers as shown on Figure 2-3.

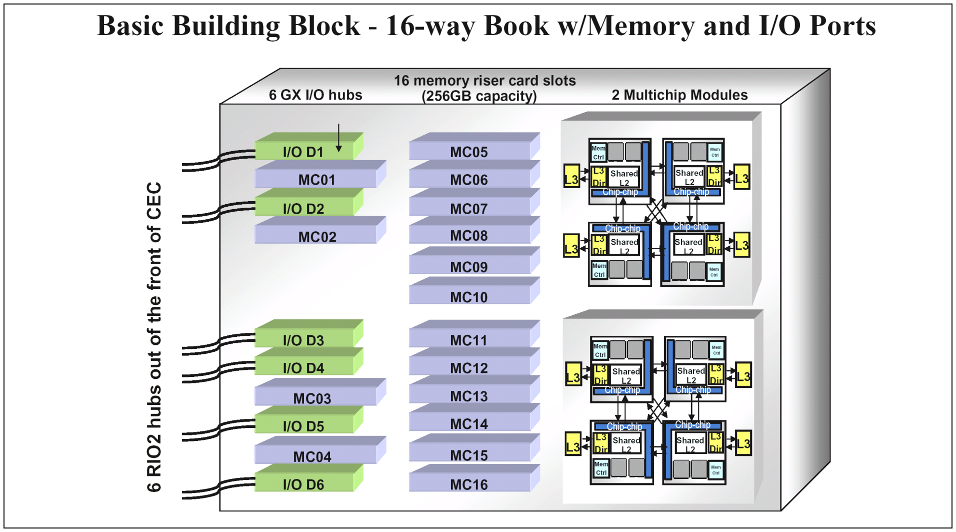


Figure 2-3 p5-590 and p5-595 16-way processor book diagram

2.2.6 Processor clock rate

The p5-590 system features base 8-way (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 and 1.9 GHz.

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

To determine the processor characteristics on a running system, use one of the following commands:

lsattr -El proc*X* Where *X* is the number of the processor; for example, proc0 is the first processor in the system. The output from the command[3] would be similar to this:

```
type powerPC_POWER5      Processor type        False
frequency 165600000      Processor Speed       False
smt_enabled true         Processor SMT enabled False
smt_threads 2            Processor SMT threads False
state enable             Processor state       False
```

(False, as used in this output, signifies that the value cannot be changed through an AIX command interface.)

pmcycles -m This command (AIX 5L Version 5.3 and later) uses the performance monitor cycle counter and the processor real-time clock to measure the actual processor clock speed in MHz. This is the sample output of a 8-way p5-590 running at 1.65 GHz system:

```
Cpu 0 runs at 1656 MHz
Cpu 1 runs at 1656 MHz
Cpu 2 runs at 1656 MHz
Cpu 3 runs at 1656 MHz
Cpu 4 runs at 1656 MHz
Cpu 5 runs at 1656 MHz
Cpu 6 runs at 1656 MHz
Cpu 7 runs at 1656 MHz
```

> **Note:** The **pmcycles** command is part of the bos.pmapi fileset. First check whether that component is installed by using the **lslpp -l bos.pmapi** command in AIX.

[3] The output of the **lsattr** command has been expanded with AIX 5L to include the processor clock rate.

2.3 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 chips as shown on Figure 2-4. There are 16 memory card slots per processor book and each processor chip on an MCM owns a pair of memory cards. The GX interface provides I/O subsystem connection.

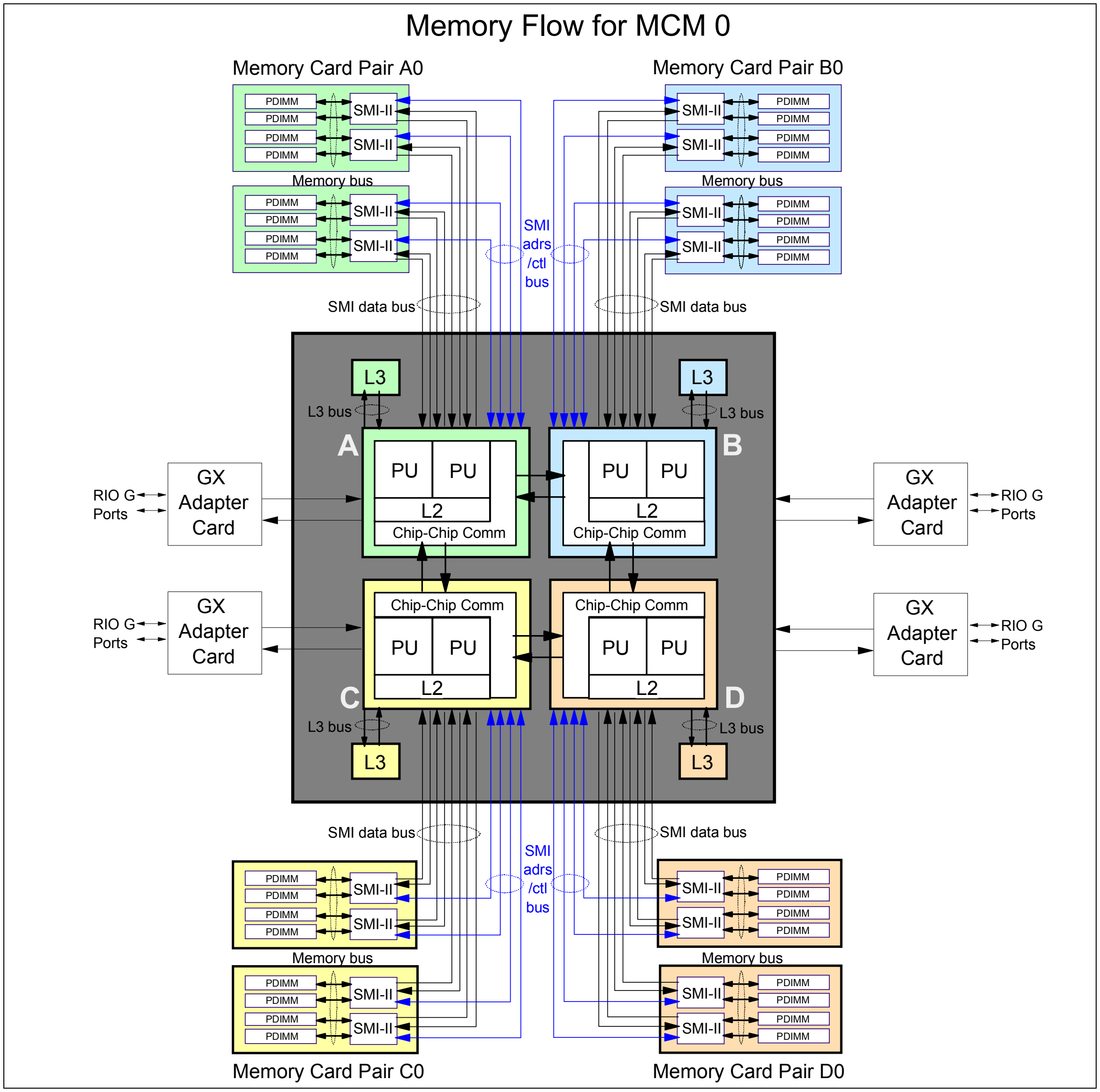


Figure 2-4 Memory flow diagram for MCM0

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 processor cards. Table 2-1 lists the possible memory configurations.

Table 2-1 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[a]	32	64

[a] Number of installable memory cards depends on number of installed processor books. (16 per processor book) Memory cards are installed in quads.

2.3.1 Memory cards

On the p5-590 and the p5-595 systems the memory is seated on a memory card shown on Figure 2-5. Each memory card has four soldered DIMM cards and two SMI-II chips for address/controls, and data buffers. Individual DIMM cards cannot be removed or added and memory cards have fixed amount of memory. Table 2-2 on page 28 lists the available type of memory cards.

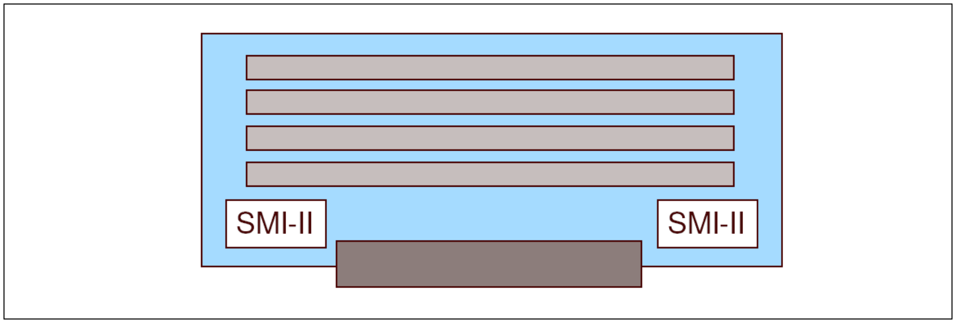


Figure 2-5 Memory card with four DIMM slots

2.3.2 Memory placement rules

The memory features that are available for the p5-590 and the p5-595 at the time of writing are listed in Table 2-2. The memory locations for each processor chip in the MCMs are illustrated in Figure 2-6 on page 28.

> **Note:** DDR1 and DDR2 cannot be mixed within a p5-590/p5-595 server.

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

Memory type	Size	Speed	Number of memory cards	Feature code
DDR1 COD	4 GB (2 GB active)	266 MHz	1	7816
	8 GB (4 GB active)	266 MHz	1	7835
DDR1	16 GB	266 MHz	1	7828
	32 GB	200 MHz	1	7829
	256 GB package	266 MHz	32 * 8 GB	8195
	512 GB package	266 MHz	32 * 16 GB	8197
	512 GB package	200 MHz	16 * 32 GB	8198
DDR2	4 GB	533 MHz	1	7813

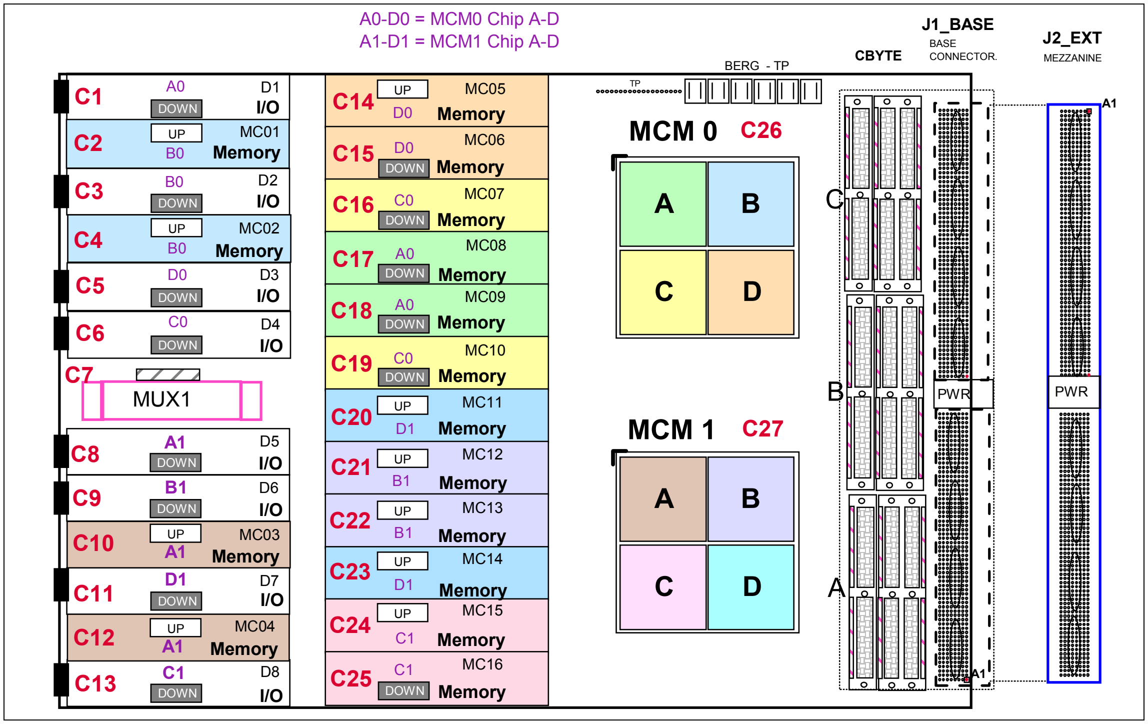


Figure 2-6 Memory placement for the p5-590 and p5-595

The following memory configuration guidelines are recommended.

- p5-590/p5-595 servers with one processor book must have a minimum of two memory cards installed.
- Each 8-way MCM (two per processor book) should have memory installed.
- The same amount of memory should be used for each MCM (two per processor book) in the system.
- No more than two different sizes of memory cards should be used in any processor book.
- All MCMs (two per processor book) in the system should have the same total 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 high-performance computing the following in strongly recommended.

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

2.4 Central electronics complex

The Central Electronics Complex (CEC) is an 18 EIA unit drawer that houses the processors and memory cards of the p5-590 and p5-595. The fundamental differences between p5-590 and p5-595 are outlined in Section 1.3, “General overview and characteristics” on page 4.

The CEC contains the following components:

- CEC backplane that serves as the system component mounting unit
- Multichip modules (MCMs) books that contains the POWER5 processors and L3 cache modules
- Memory cards
- Service processor unit
- I/O books that provide the Remote I/O (RIO) ports
- Fans and blowers for CEC cooling

Figure 2-7 on page 31 provides a logical view of the CEC components. It shows a system populated with two MCMs, eight dual core POWER5 processors, eight L3 cache modules, eight GX+ adapter card slots, the memory subsystem, and dual processor clock.

The CEC houses up to four processor books. Each processor book consists of two MCMs, memory, and GX+ ports (for Remote I/O or for HPS). A single MCM has exactly four dual-core POWER5 chips and their Level 3 cache memory (4 x 36 MB). A p5-595 supports maximum of eight MCMs and is 64-way (4 x 2 x 8).

The p5-595 supports a maximum configuration of four processor books, which allows for 64-way processing.

The p5-590 is packaged in the same physical structure as the p5-595. However, the p5-590 supports a maximum of two processor books, which allows for 32-way processing.

The enhanced GX+ high frequency processor bus drives synchronous memory Interface (SMI) memory buffer chips and interfaces with a 36 MB L3 cache controller. The processor bus provides eight bus adapter slots to interface with two types of bus technologies:

- I/O bus bridges
- CEC interconnect bridge to support clustered system configurations. The p5-590 does not support clustered configurations.

The main service processor function is located in the CEC. The service processor subsystem runs an independent operating system and directly drives Ethernet ports to connect the external hardware management console (HMC). The service processor unit is not specifically represented in the CEC logic diagram, Figure 2-7 on page 31. The service processor is explained in more detail in Chapter 7, “Service processor” on page 169.

There are 16 memory card slots available for the p5-595 and p5-590. For detailed information about memory, refer to Section 2.3.2, “Memory placement rules” on page 27.

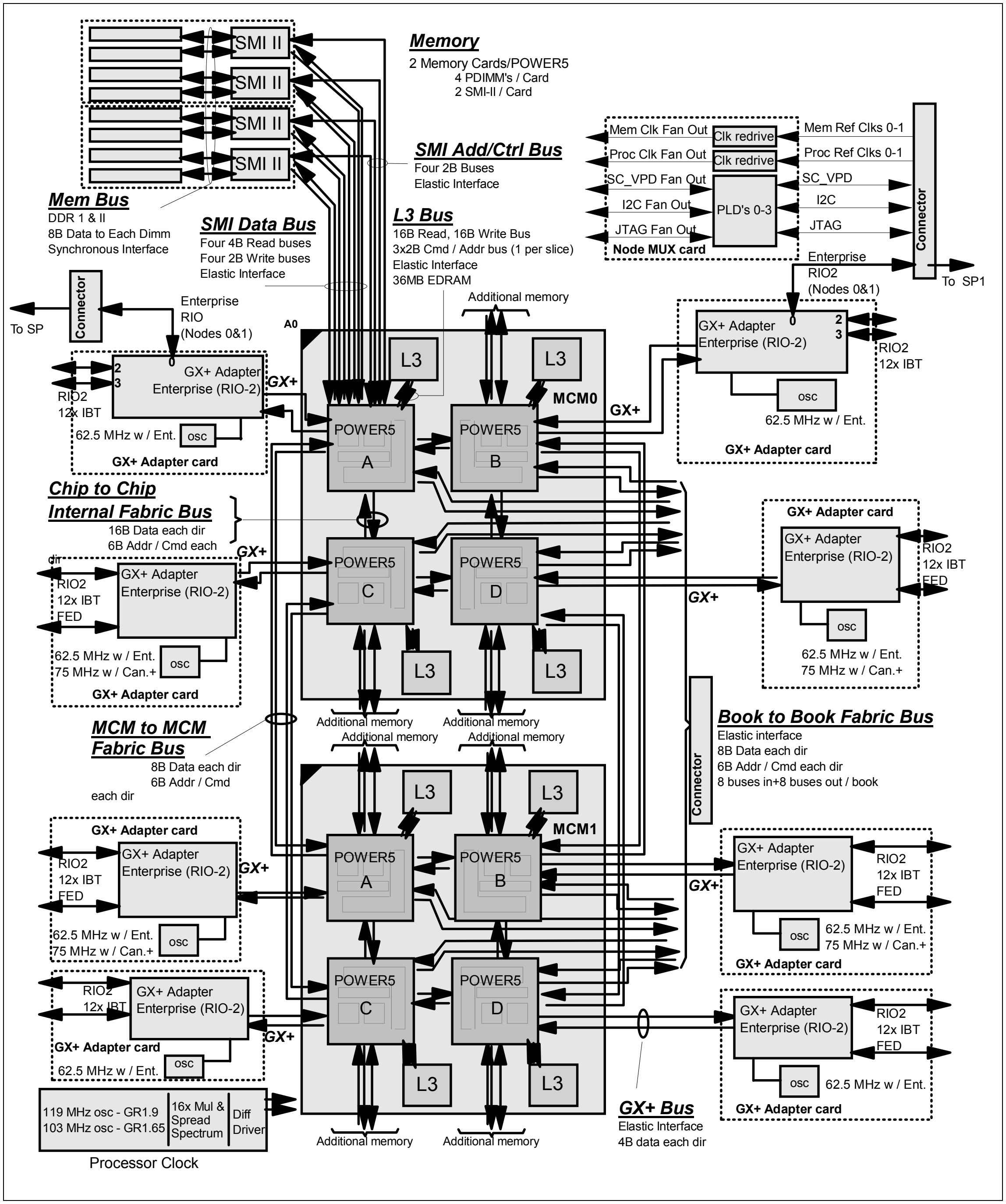


Figure 2-7 p5-595 and p5-590 CEC logic diagram

Major design efforts have contributed to the development of the p5-590 and p5-595 to analyze single points of failure within the CEC and to either eliminate them or to provide hardening capabilities to significantly reduce their probability of failure.

2.4.1 CEC backplane

The CEC backplane is a double-sided passive backplane that serves as the mounting unit for various system components. The top view of p5-595 CEC is shown in Figure 2-8; the p5-590 CEC is show in Figure 2-9 on page 32. There are no physical differences between the p5-590 backplane and the p5-595 backplane.

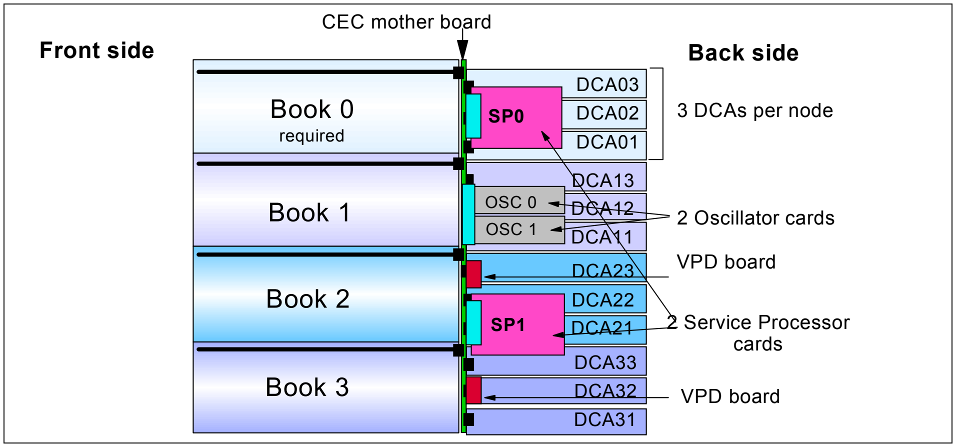


Figure 2-8 p5-595 CEC (top view)

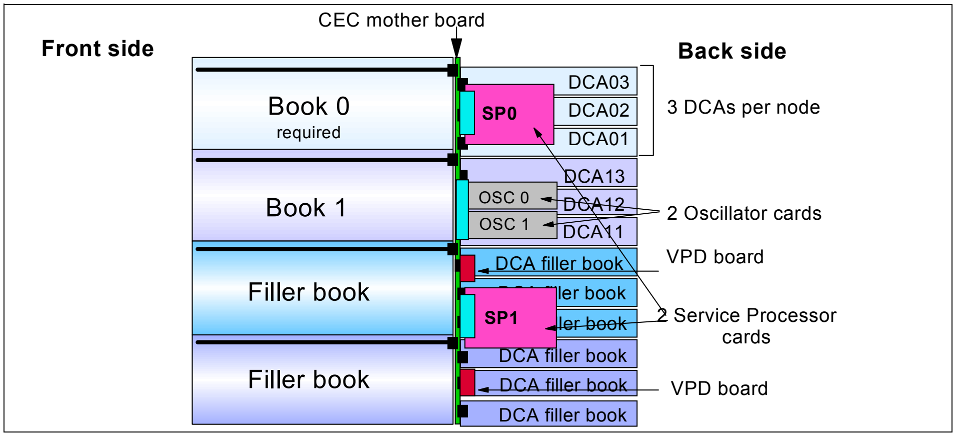


Figure 2-9 p5-590 CEC (top view)

The backplane is positioned vertically in the center of the CEC, and provides mount spaces for processor books with sixteen POWER5 processors and sixteen level 3 (L3) cache modules, eight memory cards, and four I/O books. Figure 2-10 on page 33 depicts component population on both the front side and back side of the backplane.

The clock card distributes sixteen paired clock signals to the four processor books. The connection scheme for each processor book consists of a pair of parallel connectors, a base and a mezzanine. As seen in Figure 2-10 on page 33, there are twelve distributed converter assembly (DCA) connectors, two service processor card connectors and two clock card connectors on the back side of the backplane. In addition to the logic card connections there are bus bar connectors on the front side on the top and bottom of the card.

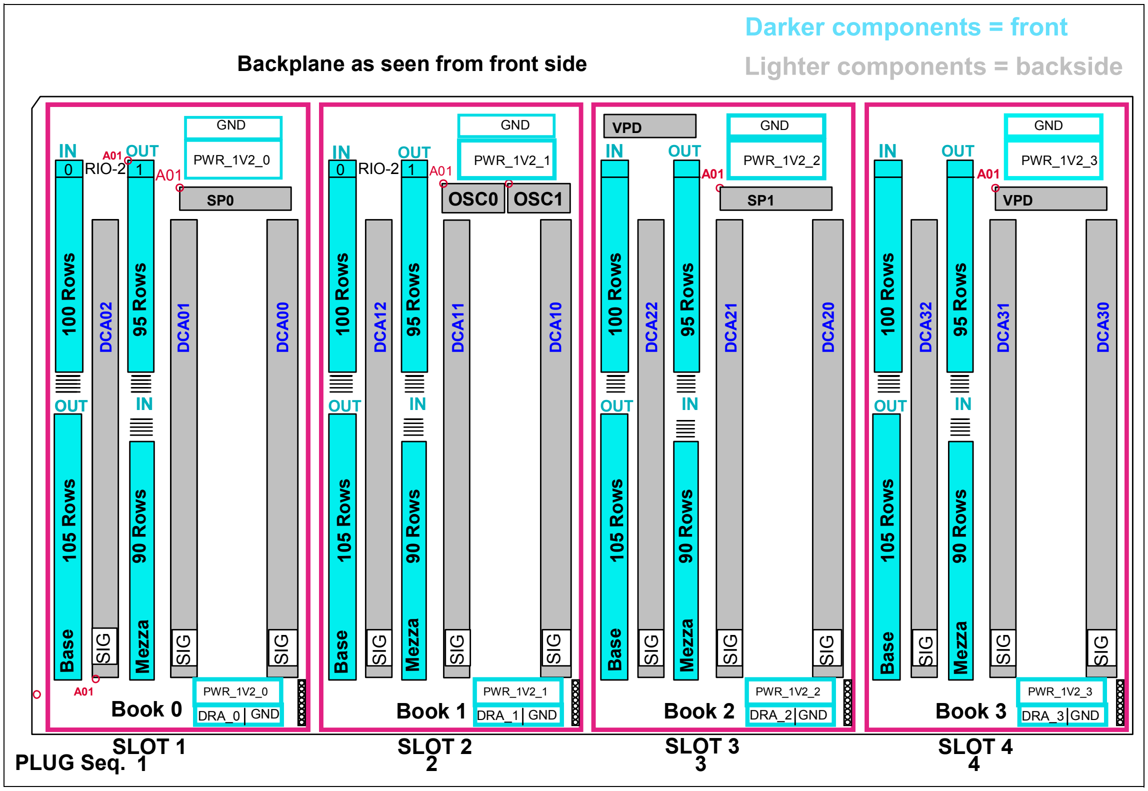


Figure 2-10 CEC backplane (front side view)

The CEC backplane provides the following types of slots:

- Slots for DCA books and two capacitor books

These slots are populated by up to six DCA books and up to two capacitor books. Each DCA book contains two DCA connectors. They supply electricity power to the CEC backplane and convert voltage.

- GX bus slots 0-3

  The GX bus slot 0 is used to insert the primary I/O book. The GX bus slots 1, 2, and 3 are used for the optional secondary I/O books. The p5-590 is restricted to at most one secondary I/O book, while the p5-595 can have from zero to three secondary books. Due to system thermal limitations, two of the slots in the air flow path of MCM0 will not be populated with GX+ adapters. These slots will be reserved to allow a fresh air path to MCM0 for cooling. Refer to Figure 2-7 on page 31 to observe the four GX+ adapter slots corresponding to MCM0.

- GX+ Bus Adapter Card Slots

  The CEC has eight GX+ slots which support communication with two types of bus technologies.

- Remote I/O (RIO-2) links allow for connectivity to external I/O drawers and PCI-X technology
- Switches to allow connectivity with another CEC to create a clustered configuration

2.5 System flash memory configuration

In the p5-590 and p5-595, a serial electronically erasable programmable read only memory (sEEPROM) adapter plugs into the back of the central electronics complex backplane. The platform firmware binary image is programmed into the system sEEPROM, also known as *system FLASH memory*. FLASH memory is initially programmed during manufacturing of the p5-590 and p5-595 systems. However, this single binary image can be reprogrammed to accommodate firmware fixes provided to the client using the hardware management console (HMC).

The firmware binary image contains boot code for the p5-590 and p5-595. This boot code includes, but is not limited to, system service processor code, code to initialize the POWER5 processors, memory, and I/O subsystem components, partition management code, and code to support the Advanced POWER Virtualization feature. The firmware binary image also contains hardware monitoring code used during partition run time.

During boot time, the system service processor dynamically allocates the firmware image from flash memory into main system memory. The firmware code is also responsible for loading the operating system image into main memory.

Additional information regarding the system service processor can be found in Chapter 7, “Service processor” on page 169.

Refer to section Section 7.4, “Firmware updates” on page 186 for a summary of the firmware update process. Refer to Section 2.4, “Central electronics complex” on page 29 for more information regarding the system CEC.

2.6 Vital product data and system smart chips

Vital product data (VPD) carries all of the necessary information for the service processor to determine if the hardware is compatible and how to configure the hardware and chips on the card. The VPD also contains the part number and serial number of the card used for servicing the machine as well as the location information of each device for failure analysis. Since the VPD in the card carries all information necessary to configure the card, no card device drivers or special code has to be sent with each card for installation.

Smart chips are micro-controllers used to store vital product data (VPD). The smart chip provides a means for securely storing data that can not be read, altered, or written other than by IBM privileged code. The smart chip provides a means of verifying IBM @server On/Off Capacity on Demand (CoD) and IBM @server Capacity Upgrade on Demand activation codes that only the smart chip on the intended system can verify. This allows clients to purchase additional spare capacity and pay for use only when needed. The smart chip is the basis for the CoD function and verifying the data integrity of the data stored in the card.

2.7 I/O drawer

The p5-590 and p5-595 use remote I/O drawers (that are 4 EIA-high) for directly attached PCI or PCI-X adapters and SCSI disk capabilities. Each I/O drawer is divided into two separate 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-swappable disk bays per drawer.

A minimum of one I/O drawer (FC 5791 or FC 5794) is required per system. I/O drawer feature number 5791 contains 20 PCI-X slots and 16 disk bays, and feature number 5794 contains 20 PCI-X slots and 8 disk bays.

Existing 7040-61D I/O drawers may be attached to a p5-590 or p5-595 as additional I/O drawers, if available. Only 7040-61D I/O drawers containing feature number 6571 PCI-X planars are supported. FC 6563 PCI planars must be replaced with FC 6571 PCI-X planars before the drawer can be attached. RIO-2 drawer interconnects is the only supported protocol (as opposed to the older RIO) in the p5-590 and p5-595.

Only adapters supported on the p5-590 and p5-595 feature I/O drawers are supported in 7040-61D I/O drawers, if attached. Unsupported adapters must be removed before attaching the drawer to the p-590 and p5-595 server. The p5-590 and p5-595 only support EEH adapters when partitioned.

A maximum of eight I/O drawers can be connected to a p5-590. Each I/O drawer contains twenty 3.3-volt PCI-X adapter slots and up to sixteen disk bays. Fully configured, the p5-590 can support 160 PCI adapters and 128 disks at 15,000 RPM.

A maximum of 12 I/O drawers can be connected to a p5-595. Each I/O drawer contains twenty 3.3-volt PCI-X adapter slots and up to sixteen disk bays. Fully configured, the p5-595 can support 240 PCI adapters and 192 disks at 15,000 RPM.

A blind-swap cassette (equivalent to those in FC 4599) is provided in each PCI-X slot of the I/O drawer. Cassettes not containing an adapter will be shipped with a plastic filler card installed to help ensure proper environmental characteristics for the drawer. If additional blind-swap cassettes are needed, FC 4599 should be ordered.

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 blind-swap backplanes in that half of the drawer and these adapters do not support external SCSI device attachments. Each half of the I/O drawer is powered separately. FC 5791 is a 7040-61D with 16 disk bays and FC 5794 is a 7040-61D with eight disk bays.

2.7.1 EEH adapters and partitioning

The p5-590 and p5-595 systems are currently only orderable with adapters that support EEH. Support of a non-EEH adapter (OEM adapter) is only possible when the system has not been configured for partitioning. This is the case when a new system is received, for example, and it is in full system partition and is planned to be used without an HMC. EEH will be disabled for that adapter upon system initialization.

When the platform is prepared for partitioning or is partitioned the POWER Hypervisor prevents disabling EEH upon system initialization. Firmware in the partition will detect any non-EEH device driver that are installed and not configure them. Therefore, all adapters in p5 systems must be EEH capable in order to be used by a partition. This applies to I/O installed in I/O drawers attached to a p5 system.

A client does not need to actually create more than a single partition to put the platform in a state where the hypervisor considers it to be partitioned. The platform becomes partitioned (in general, but also in specific reference to EEH enabled by default) as soon as the client attaches an HMC and performs any function that relates to partitioning. Simple hardware service operations do not partition the platform, so it is not simply connecting an HMC that has this affect. But, modifying any platform attributes related to partitioning (such as booting under HMC control to only PHYP standby, and suppressing autoboot to the pre-installed OS partition) results in a partitioned platform, even if the client does not actually create additional partitions.

All p5 platform IO slots are managed the same with respect to EEH.

2.7.2 I/O drawer attachment

System I/O drawers are connected to the p5-590 and p5-595 CEC using RIO-2 loops. Drawer connections are made in loops to help protect against a single point-of-failure resulting from an open, missing, or disconnected cable. Systems with non-looped configurations could experience degraded performance and serviceability. The system has a non-looped configuration if only one RIO-2 path is running.

RIO-2 loop connections operate at 1 GHz. RIO-2 loops connect to the system CEC using RIO-2 loop attachment adapters (FC 7818). Each of these adapters has two ports and can support one RIO-2 loop. Up to six of the adapters can be installed in each 16-way processor book. Up to 8 or 12 I/O drawers can be attached to the p5-590 or p5-595, depending on the model and attachment configuration.

I/O drawers may be connected to the CEC in either single-loop or dual-loop mode. Dual-loop mode is recommended whenever possible as it provides the maximum bandwidth between the I/O drawer and the CEC.

- Single-loop (Figure 2-11) mode connects an entire I/O drawer to the CEC using one RIO-2 loop (2 ports). The two I/O planars in the I/O drawer are connected together using a short RIO-2 cable. Single-loop connection requires one loop (2 ports) per I/O drawer.
- Dual-loop (Figure 2-12) mode connects each I/O planar in the drawer to the CEC separately. Each I/O planar is connected to the CEC using a separate

RIO-2 loop. Dual-loop connection requires two loops (4 ports) per I/O drawer. With dual-loop configuration, the RIO-2 bandwidth for the I/O drawer is higher.

Table 2-3 lists the number of single-looped and double-looped I/O drawers that can be connected to a p5-590 or p5-595 server based on the number of processor books installed:

Table 2-3 Number of possible I/O loop connections

Number of processor books	Single-looped	Dual-looped
1	6	3
2	8 (590) 12 (595)	6
3	12 (p5-595)	9 (p5-595)
4	12 (p5-595)	12 (p5-595)

On initial orders of p5-590 or p5-595 servers, IBM manufacturing will place dual-loop-connected I/O drawers as the lowest numerically designated drawers followed by any single-looped I/O drawers.

- A minimum of two cables are required for each loop for each GX adapter. Interconnection between drawers in a loop requires a additional RIO cable.
- FC 7924 (0.6 m) can only be used as a jumper cable to connect the two I/O drawer planars in a single loop.
- FC 3147 (3.5 m) can only be used to connect FC 5791/5794 risers that are in either a FC 5792 rack or the FC 8691 rack bolted to the primary rack to the GX adapters in a processor book.
- For the 9119-595 FC 3170 (8.0 m) can only be used to connect FC 5791/5794 risers that are in either a FC 5792 rack or the FC 8691 rack bolted to the FC 5792 rack to the GX adapters in a processor book.
- I/O drawer RIO-2 61D are limited to RIO-2 cables no longer than 8 m.
- For GX adapters to I/O drawer cabling, the first I/O drawer is connected to the first GX adapter. The second I/O drawer to the next available GX adapter. All double-looped drawers will be connected first and then the single-looped.
- The RIO cabling is GX adapter port 0 to I/O drawer riser port 0 and GX adapter port 1to I/O drawer port 1.

2.7.3 Full-drawer cabling

For an I/O drawer, the following connections are required.

- One cable from the P1 RIO Riser card J0 to CEC I/O card Px-Cx-T1.

- One cable from the P2 RIO Riser card J1 to CEC I/O card Px-Cx-T2.

  These cables provides a data and communication path between the memory cards and the I/O drawers (Figure 2-11 on page 39).

- A cable is also added between P1 RIO Riser card J1 and P2 RIO Riser card J0 in each drawer.

  This cable ensures that each side of the drawer (P1 and P2) can be accessed by the CEC I/O (RIO-2 adapter) card, even if one of the cables are damaged. Each half of the I/O drawer can communicate with the CEC I/O card for its own uses or on behalf of the other side of the drawer.

There is an identifier (ID) in the cable plugs which gives the length of the cable to the Inter-Integrated Circuit (I2C) bus and the service processor.

The cable ID is the function that verifies the length of the RIO-2 cable. There are different RIO-2 cables, because we use CEC frame, powered 24 inch A frame, and unpowered 24 inch Z frame for the I/O Drawer. With the cable ID we calculate the length and the link speed for the RIO-2 cable.

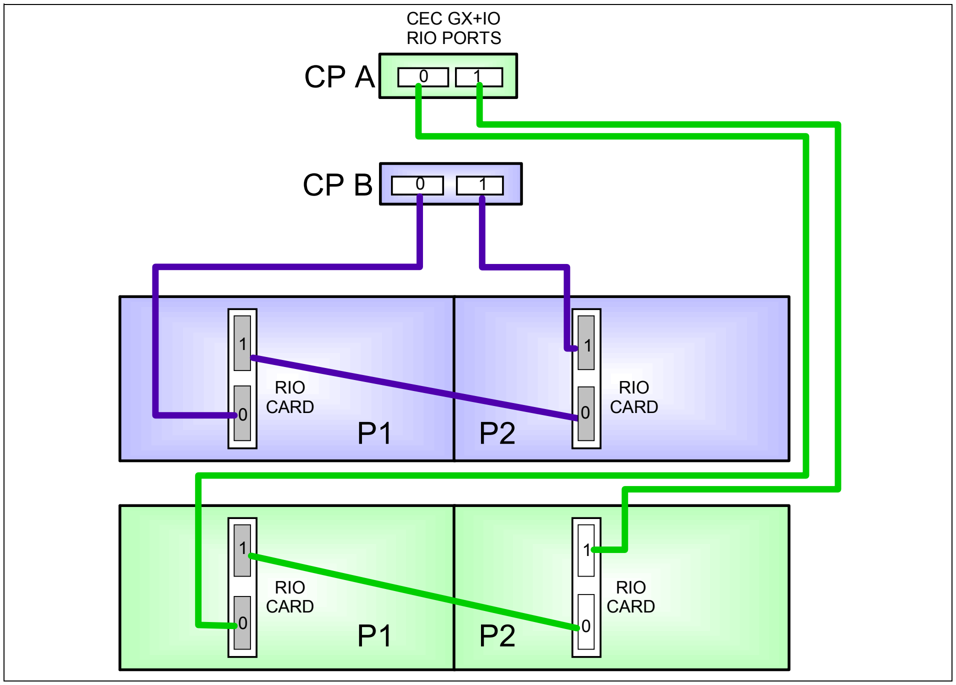


Figure 2-11 Single loop 7040-61D

2.7.4 Half-drawer cabling

Although I/O drawers will not be built in half-drawer configurations, they can be cabled to, and addressed by the CEC, in half drawer increments (Appendix 2-12, “Dual loop 7040-61D” on page 40).

Both STI connectors on one CEC I/O card Px-Cx-T1 and Px-Cx-T2 will be cabled to both ports on P1 RIO Riser card.

Both STI connectors on a different CEC I/O card Px-Cx-T1 and Px-Cx-T2 (possibly on a different processor book) will be cabled to both ports on P2 RIO Riser card.

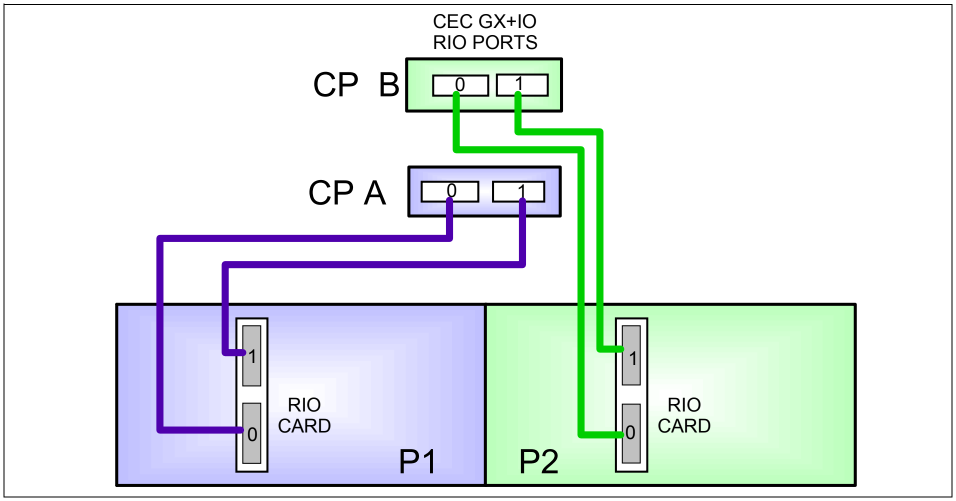


Figure 2-12 Dual loop 7040-61D

However, to simplify the management of the server we strongly recommend that I/O loops be configured as described in the IBM @server Information Center, and to only follow a different order when absolutely necessary.

In any case, it becomes extremely important for the management of the system to keep an up-to-date cabling documentation of your systems, because it may be different from the cabling diagrams of the installation guides.

The I/O drawers provide internal storage and I/O connectivity to the system. Figure 2-13 shows the rear view of an I/O drawer, with the PCI slots and riser cards that connect to the RIO ports in the I/O books.

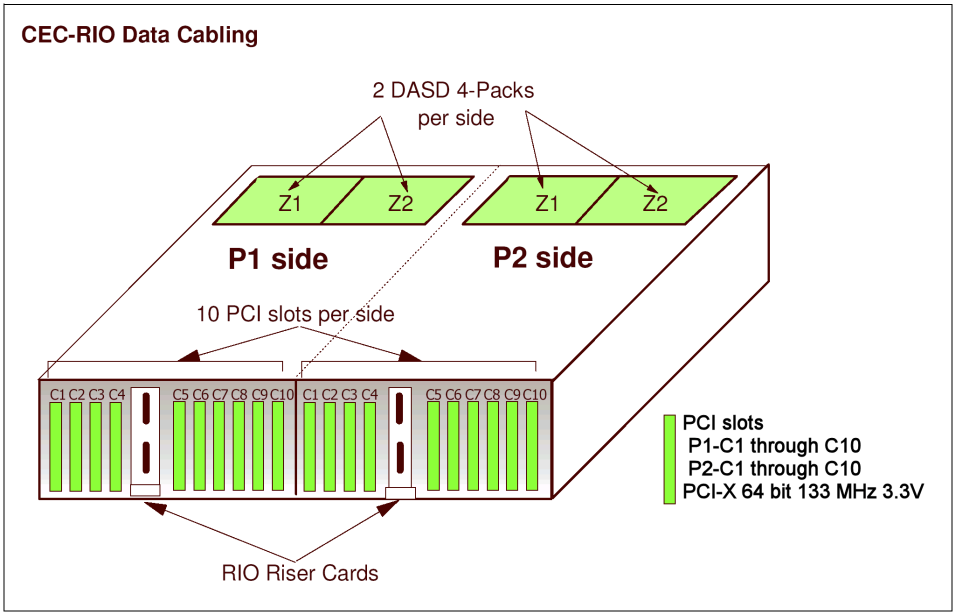


Figure 2-13 I/O drawer -GX+

Each drawer is composed of two physically symmetrical I/O planar boards that contain 10 Hot-Plug PCI-X slots each, and PCI adapters can be inserted in the rear of the I/O drawer. The planar boards also contain two integrated Ultra3 SCSI adapters and SCSI Enclosure Services (SES), connected to a SCSI 4-pack backplane.

2.7.5 Blind-swap cassette

Also named the blind-swap mechanism (Figure 2-14) or PCI carrier, each PCI (I/O) Card must be housed in a blind-swap cassette before being installed.

All PCI slots on the p5-590 and p5-595 are PCI 2.2-compliant and are Hot-Plug enabled, which allows most PCI adapters to be removed, added, or replaced without powering down the system. This function enhances system availability and serviceability.

The function of hot-plug PCI adapters is to provide concurrent adding or removal of PCI adapters when the system is running. In the I/O drawer, the installed adapters are protected by plastic separators called blind-swap cassettes. These are used to prevent grounding and damage when adding or removing adapters. The Hot-Plug LEDs outside the I/O drawer indicate whether an adapter can be plugged into or removed from the system.

The Hot-Plug PCI adapters are secured with retainer clips on top of the slots; therefore, you do not need a screwdriver to add or remove a card, and there is no screw that can be dropped inside the drawer.

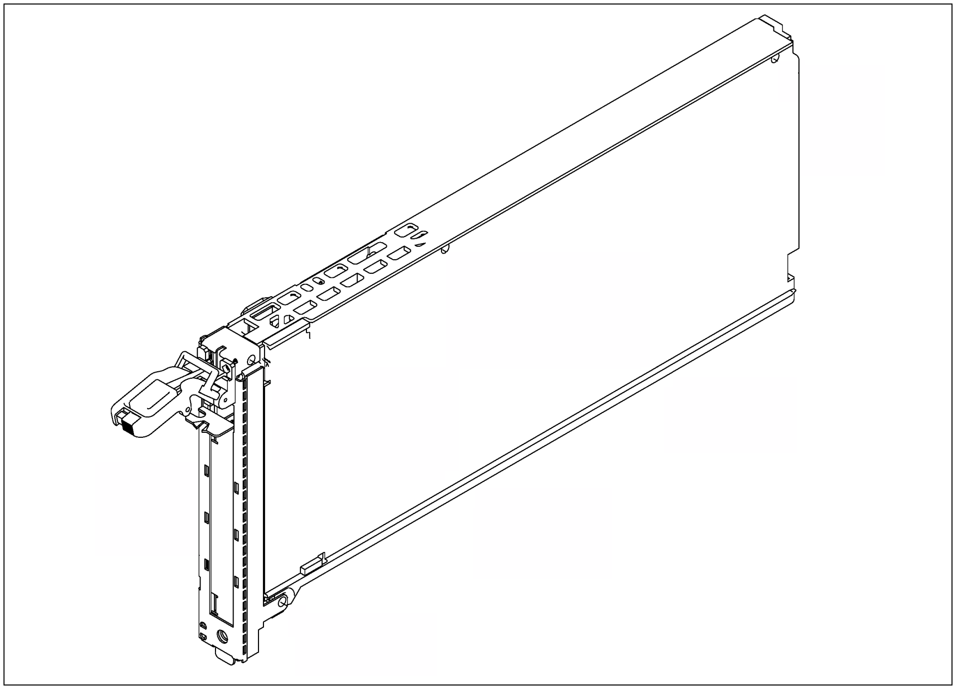


Figure 2-14 Blind-swap cassette

2.7.6 Logical view of a RIO-2 drawer

Figure 2-15 shows a logical schematic of an I/O drawer and the relationship between the internal controllers, disks, and I/O slots.

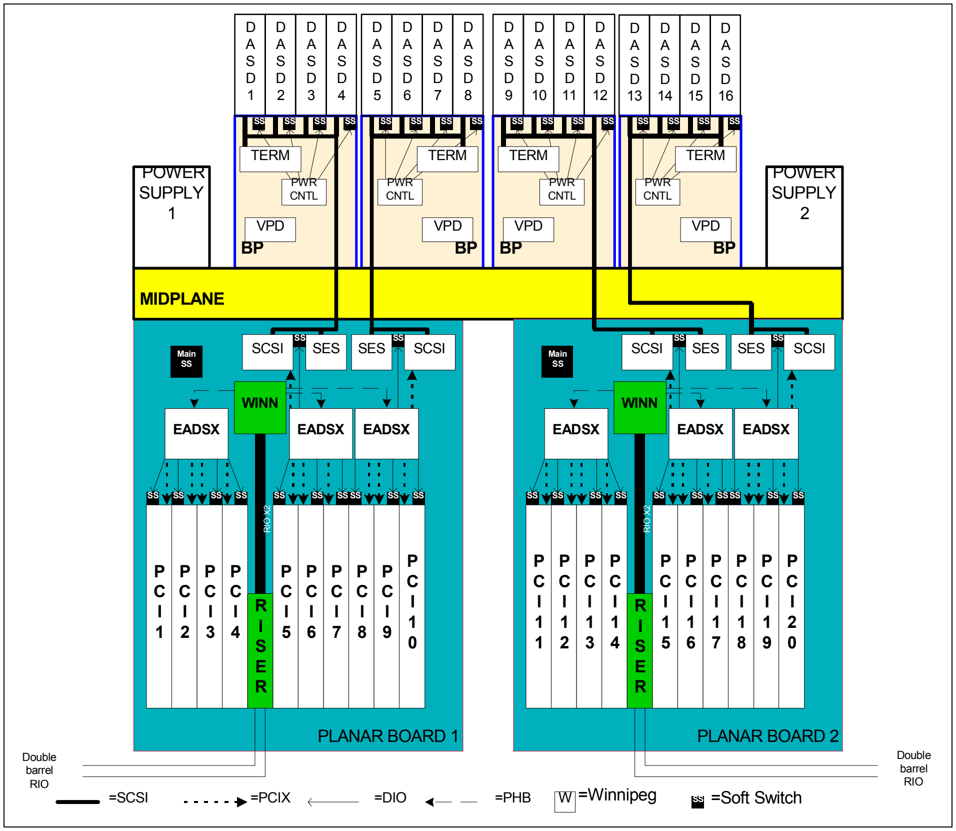


Figure 2-15 I/O drawer top view - logical layout

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

The 36.4 GB and73.4 GB disks have the following characteristics:

- Form factor: 3.5-inch, 1-inch (25 mm) high
- SCSI interface: SCSI Ultra3 (Fast 80) 16 bit
- Rotational speed: 15,000 RPM (disks with 10K rotational speeds from earlier systems are not supported)

The RIO riser cards are connected to the planar boards. The RIO ports of each riser card are connected through I/O loops to RIO ports on I/O books in the CEC. The connectivity between the I/O drawer RIO ports and the I/O books RIO ports is described in Remote I/O loop.

On each planar board, the ten PCI-X slots have a 3.3V PCI bus signaling and operates at 33 MHz, 66 MHz, or 133 MHz, depending on the adapter. All PCI slots are PCI 2.2 compliant and are Hot-Plug enabled (Figure 2.7.6).

PCI adapters have different bandwidth requirements, and there are functional limitations on the number of adapters of a given type in an I/O drawer or a system.

The complete set of limitations are described in the Hardware Information Center. This is regularly updated and should be considered as the reference for any questions related to PCI limitations.

In a RIO-2 I/O-drawer all the I/O slots can be populated with high speed adapters (for example, Gigabit Ethernet, Fiber Channel, ATM or Ultra-320 SCSI adapters). All can be populated, but in some situations we might not get optimum performance on each by bandwidth limitation.

2.7.7 I/O drawer RAS

If there is an RIO failure in a port or cable, an I/O planar board can route data through the other I/O connection and share the remaining RIO cable for I/O.

For power and cooling, each drawer has two redundant DC power supplies and four high reliability fans. The power supplies and fans have redundancy built into them, and the drawer can operate with a failed power supply or a failed fan. The hard drives and the power supplies are hot-swappable, and the PCI adapters are Hot-Plug.

All power, thermal, control, and communication systems are redundant in order to eliminate outages due to single-component failures.

I/O subsystem communication and monitoring

There are two main communication subsystems between the CEC and the I/O drawers. The power and RAS infrastructure are responsible for gathering environmental information and controlling power on I/O drawers. The RIO loops are responsible for data transfer to and from I/O devices.

Power and RAS infrastructure

The power cables that connect each I/O drawer and the bulk power assembly (BPA) provide both electricity power distribution and Reliability, Availability, and Serviceability (RAS) infrastructure functions that include:

- Powering all system components up or down, when requested. These components include I/O drawers and the CEC.
- Powering down all the system enclosures on critical power faults.

- Verifying power configuration.
- Reporting power and environmental faults, as well as faults in the RAS infrastructure network itself, on operator panels and through the service processor.
- Assigning and writing location information into various VPD elements in the system.

> **Note:** It is the cabling between the RIO drawer and the BPA that defines the numbering of the I/O drawer not the physical location of the drawer.

The power and RAS infrastructure monitors power, fans, and thermal conditions in the system for problem conditions. These conditions are reported either through an interrupt mechanism (for critical faults requiring immediate operating system action) or through messages passed from the RAS infrastructure to the service processor to Run-Time Abstraction Service (RTAS).

2.7.8 Supported I/O adapters in p5-595 and p5-590

The following are configuration rules for the I/O drawer.

I/O Drawer (5791/5794) adapters placement (p5-590 and p5-595 only)

The FC 5791 drawer provides 20 blind-swap PCI-X slots and 4 integrated DASD backplanes that support up to 16 hot-swap disk bays. The FC 5794 drawer is the same as FC 5791 but supports only two integrated DASD backplanes that support up to eight hot-swap disk bays. The 20 PCI-X slots are divided into six PCI Host Bridges (PHB) as follows:

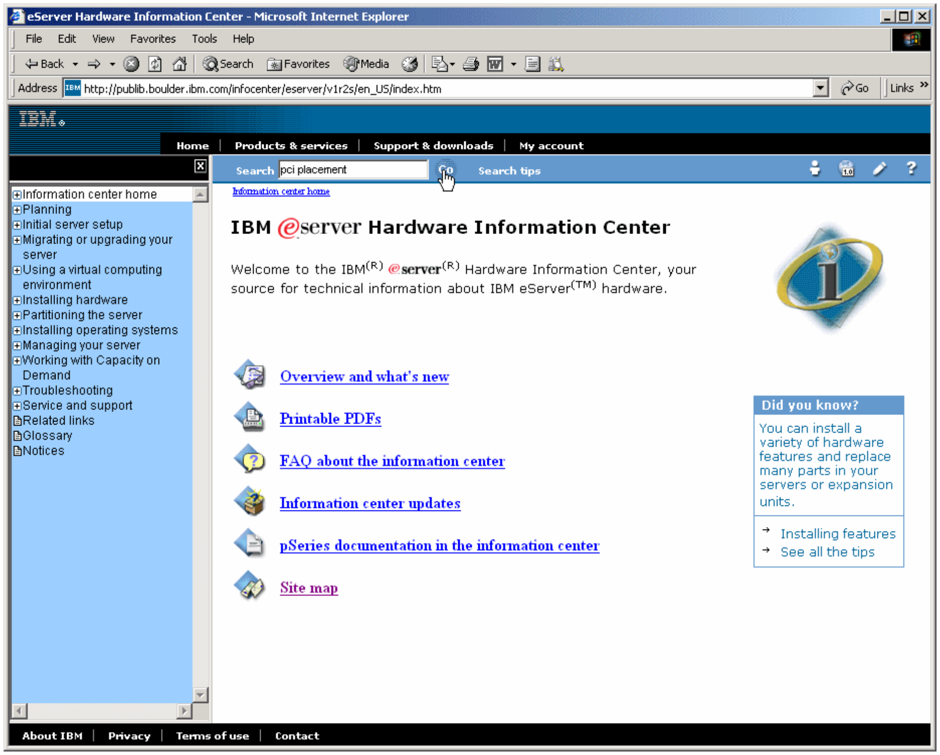
- PHB1 = slots 1, 2, 3, 4
- PHB2 = slots 5,6,7; Z1 onboard
- PHB3 = slots 8,9,10, Z2 onboard
- PHB4 = slots 11, 12, 13, 14
- PHB5 = slots 15, 16, 17, Z1 onboard
- PHB6 = slots 18, 19, 20, Z2 onboard

Figure 2-16 and Figure 2-17 shows how to find more information on PCI adapter placement in IBM @server Hardware Information Center by placing a search for *PCI placement*.

http://publib.boulder.ibm.com/infocenter/eserver/v1r2s/en_US/index.htm

and search for *pci placement* or:

`http://publib.boulder.ibm.com/infocenter/eserver/v1r2s/en_US/info/iphak/expansion61d.htm#expansion61d`


eServer Hardware Information Center - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://publib.boulder.ibm.com/infocenter/eserver/v1r2s/en_US/index.htm
Go Links
IBM
Home | Products & services | Support & downloads | My account
Search pci placement Go Search tips
About IBM | Privacy | Terms of use | Contact

- Information center home
- Planning
- Initial server setup
- Migrating or upgrading your server
- Using a virtual computing environment
- Installing hardware
- Partitioning the server
- Installing operating systems
- Managing your server
- Working with Capacity on Demand
- Troubleshooting
- Service and support
- Related links
- Glossary
- Notices

Information center home

IBM @server Hardware Information Center

Welcome to the IBM(R) @server(R) Hardware Information Center, your source for technical information about IBM eServer(TM) hardware.

- Overview and what's new
- Printable PDFs
- FAQ about the information center
- Information center updates
- pSeries documentation in the information center
- Site map

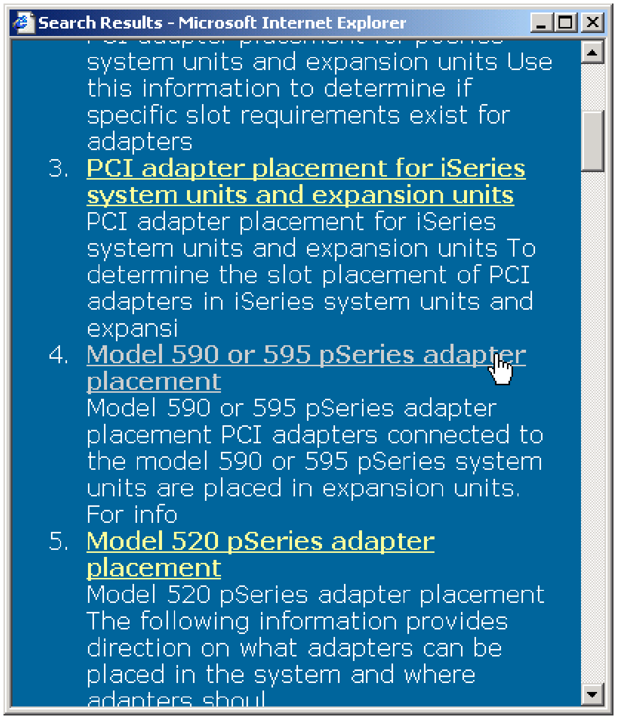
Did you know?

You can install a variety of hardware features and replace many parts in your servers or expansion units.

- Installing features
- See all the tips

Figure 2-16 Hardware Information Center search for PCI placement

Figure 2-17 on page 47 shows sample results after searching for *PCI placement*.



PCI adapter placement for pSeries system units and expansion units Use this information to determine if specific slot requirements exist for adapters

3. PCI adapter placement for iSeries system units and expansion units
   PCI adapter placement for iSeries system units and expansion units To determine the slot placement of PCI adapters in iSeries system units and expansi
4. Model 590 or 595 pSeries adapter placement
   Model 590 or 595 pSeries adapter placement PCI adapters connected to the model 590 or 595 pSeries system units are placed in expansion units. For info
5. Model 520 pSeries adapter placement
   Model 520 pSeries adapter placement The following information provides direction on what adapters can be placed in the system and where adapters shoul

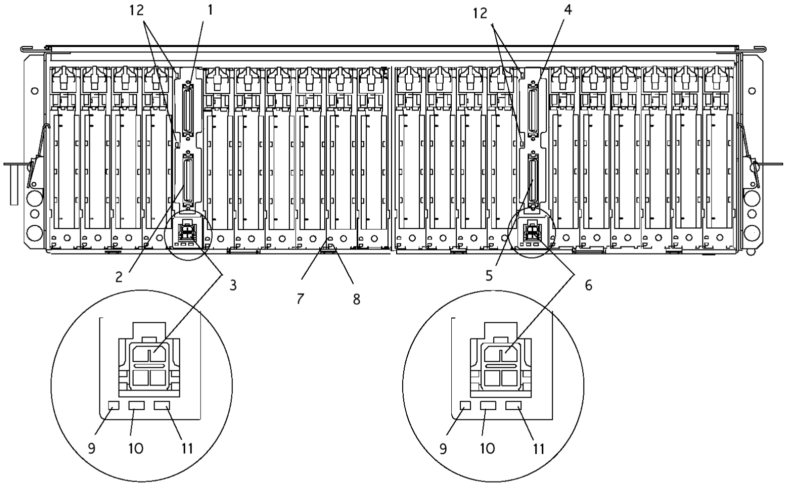
Figure 2-17 Select Model 590 or 595 placement

Adapters are placed based on the highest position in the table first, into the first slot in the slot priority for that row in the table. If that slot is filled, place the card in the next available slot in the Slot Priority for that adapter. Adapters have been divided into three performance categories as required.

- The first I/O-drawer at rack position EIA 05 must have a PCI-X Dual Channel Ultra320 SCSI Adapter (FC 5712) for connection to a media device (FC 5710). It will be placed in either slot 10 or slot 20.
- A blind-swap cassette will be assigned for every adapter card on the order. Empty slots will be assigned a blind-swap cassette with a plastic filler. Additional blind-swap cassettes (FC 4599) can be ordered through your IBM sales representative.
- Actual slot numbers are stamped on planar 1 = I1 through I10 (left to right from rear). Slot numbers stamped on planar 2 are also numbered I1 through I10 and are represented as slots 11 through 20.

Adapter placement sequence

The adapters will be spread across PHBs (PHB1 = slots 1 to 4, PHB2 = slots 5 to 7, PHB3 = slots 8 to 10, PHB4 = slots 11 to 14, PHB5 = slots 15 to 17, PHB6 = slots 18 to 20) in each drawer starting with the primary drawer (EIA 5) in the following sequence (Figure 2-18).



1 I/O port connector 1
2 I/O port connector 0
3 Media subsystem power connector (Ux.y-P1-V1/Q3)
4 I/O port connector 1
5 I/O port connector 0
6 Media subsystem power connector
7 I/O Card Power On LED (Green)
8 I/O Adapter Fault/Identify LED (Amber/Bottom)
9 Auxiliary Power Good (Green)
10 I/O Subsystem Backplane Fault (Amber)
11 I/O Subsystem Backplane Power On (Green)
12 I/O LED (Currently Unused)

Figure 2-18 PCI slots of the I/O drawer (rear view)

See Appendix B, “PCI adapter placement guide” on page 253 for more information on adapter placement.

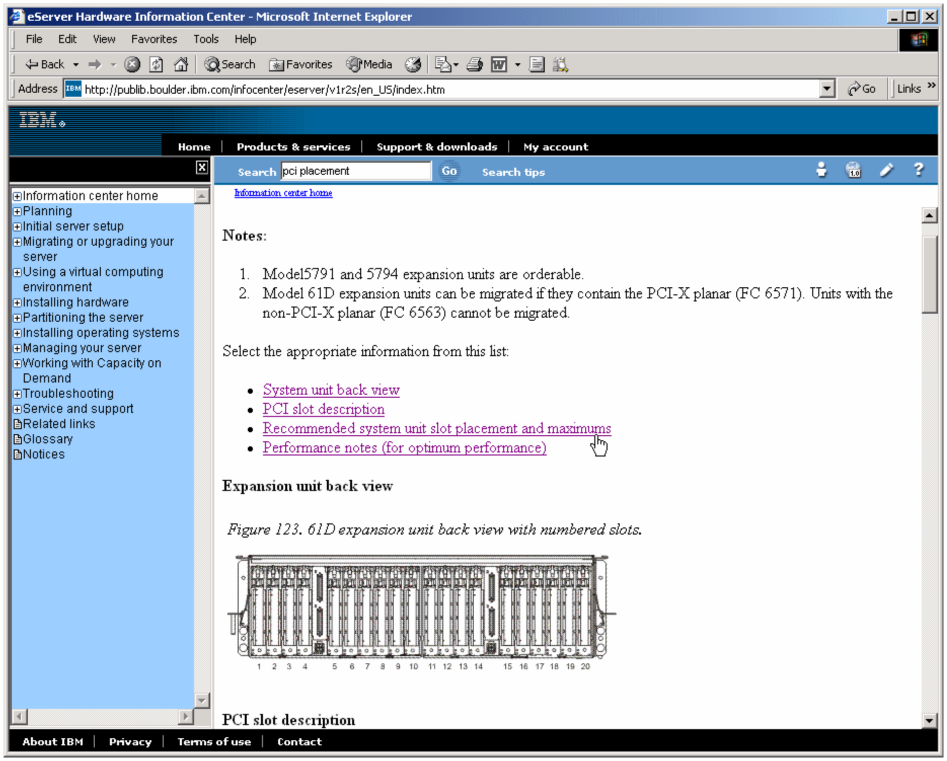
2.7.9 Expansion units 5791, 5794, and 7040-61D

The following URL for IBM @server Information Center provides direction on what adapters can be placed in the 5791, 5794, and 7040-61D expansion units and where adapters should be placed for optimum performance. Figure 2-19 shows the screen capture of the information center.

`http://publib.boulder.ibm.com/infocenter/eserver/v1r2s/en_US/index.htm`

and search for *pci placement 595* or:

http://publib.boulder.ibm.com/infocenter/eserver/v1r2s/en_US/info/iphak/expansion61d.htm#expansion61d


eServer Hardware Information Center - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://publib.boulder.ibm.com/infocenter/eserver/v1r2s/en_US/index.htm
Go
Links
IBM
Home | Products & services | Support & downloads | My account
Search pci placement Go Search tips
Information center home
Planning
Initial server setup
Migrating or upgrading your server
Using a virtual computing environment
Installing hardware
Partitioning the server
Installing operating systems
Managing your server
Working with Capacity on Demand
Troubleshooting
Service and support
Related links
Glossary
Notices
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
About IBM | Privacy | Terms of use | Contact

Information center home

Notes:

1. Model5791 and 5794 expansion units are orderable.
2. Model 61D expansion units can be migrated if they contain the PCI-X planar (FC 6571). Units with the non-PCI-X planar (FC 6563) cannot be migrated.

Select the appropriate information from this list:

- System unit back view
- PCI slot description
- Recommended system unit slot placement and maximums
- Performance notes (for optimum performance)

Expansion unit back view

Figure 123. 61D expansion unit back view with numbered slots.

PCI slot description

Figure 2-19 PCI placement guide on IBM ⓔserver information center

Model 5791 and 5794 expansion units

The following is an overview of the PCI placement information located in the Information Center. It is intended to give you an idea of what you may find there.

> **Note:** Model 7040-61D expansion units can be migrated if they contain the PCI-X planar (FC 6571). Units with the non-PCI-X planar (FC 6563) cannot be migrated.

- System unit back view
- PCI slot description
- Recommended system unit slot placement and maximums
- Performance notes (for optimum performance)

- Expansion unit back view
- PCI slot description
- Slots 1 through 20 are compatible with PCI or PCI-X adapters
- All slots support Enhanced Error Handling (EEH)
  - The Uffff.ccc.sssssss.Pn.Cm..... represents the Hardware Management Console (HMC) location code, which provides information as to the identify of the enclosure, backplane, PCI adapter(s), and connector. The ffff.ccc.sssssss in the location code represents the following:
    - ffff = Feature Code of the Enclosure (drawer or processor book)
    - ccc = the Sequence Number of the Enclosure
    - sssssss = the Serial Number of the Enclosure.

Recommended system unit slot placement and maximums

- Extra High Bandwidth (EHB) adapter. See the Performance notes before installing this adapter.
- High Bandwidth (HB) adapter. See the Performance notes before installing this adapter.
- For more information about listed adapters, see pSeries PCI and PCI-X adapters.
- System unit information
- No more than three GB Ethernet ports per PHB.
- No more than three high bandwidth adapters per PHB.
- No more than one extra high bandwidth adapter per PHB.
- No more than one 10 GB Ethernet port per two CPUs in a system. If one 10GB Ethernet port is present per two CPUs, no other 10 GB or 1 GB Ethernet ports should be installed for optimum performance.
- No more than two 1 GB Ethernet ports per one CPU in a system for maximum performance. More Ethernet adapters may be added for connectivity.

Figure 2-20 on page 51 and Figure 2-21 on page 51 shows the possible system configurations including the I/O frames.

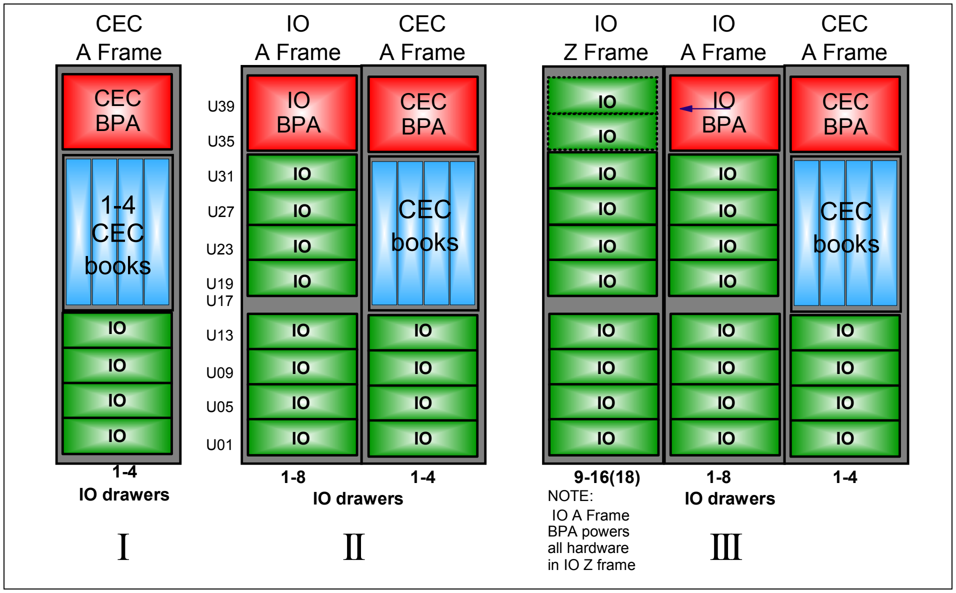


Figure 2-20 Minimum to maximum I/O configuration

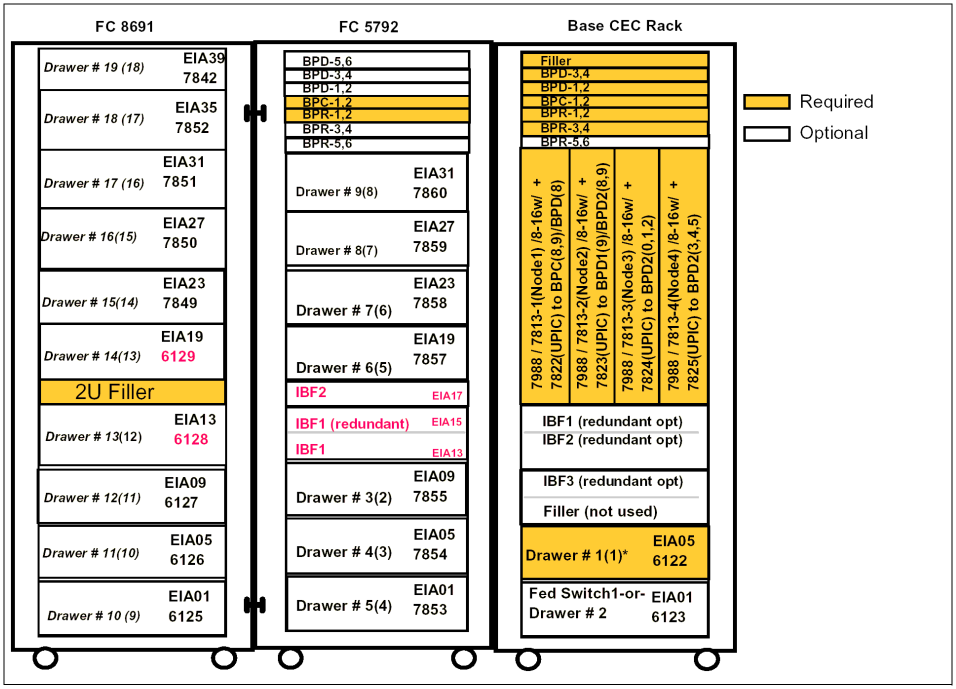


Figure 2-21 I/O frame configuration example

2.7.10 Configuration of I/O drawer ID and serial number

In some cases if the I/O Drawer was previously connected to a p650 or p655 you must configure the ID and serial number.

Using the ASMI[4] to set the configuration ID

To perform this operation, the server must be powered on and your authority level must be one of the following:

- – Administrator (Login as *Admin*)
- – Authorized service provider

- ► If the AC power is not applied, then apply it now.
- ► The drawer may power up automatically.
- ► FRU replacement will generate a new temporary unit value in the expansion unit control panel. Use this new value to power down the expansion drawer without removing the power cord using *powering off an expansion unit*. Then return here and continue with the next step.
- ► On the ASMI welcome panel, specify your user ID and password, and click Log In.
- ► In the navigation area, expand *System Configuration* and click *Configure I/O Enclosures*.
- ► Select the unit identified by the new value in the panel of the unit you are working on. In most cases it will appear as *TMPx.yyy.yyyyyyy* where x is a hex digit, and the *y's* are any value.
- ► Select *change settings*.
- ► Enter the Power Control Network Identifier:
  - – 81 for 5074 and 5079 expansion units
  - – 89 for 5088 and 0588 expansion units
  - – 8A for 5094 and 5294 expansion units
  - – 8B for 5095 and 0595 expansion units
  - – 88 for 7311-D10, 7311-D11, and 5790 expansion units
  - – 8C for 7311-D20 expansion units
- ► Enter the type-model from the label on the I/O unit.
- ► Enter the serial number (also called sequence number) from the label on the I/O unit.
- ► Click *Save Setting* to complete the operation.
- ► Do not use the browser's back button or the values will not be saved.

[4] More information on Advanced System Management Interface (ASMI) can be found in Section 7.3, "Advanced System Management Interface (ASMI)" on page 175

- Verify the correct value is now displayed in the panel of the unit you are working on.
- Disconnect all AC power to the unit, wait for the display panel to go off, and then reconnect the ac power.

> **Note:** The drawer will automatically power on. Log off and close the ASMI and return to the procedure that sent you here.

Using the control panel to set the configuration ID

To perform this operation, the server must be powered on.

Control panel function 07 is used to query and set the configuration ID and to display the frame address of any drawer connected to the SPCN network. Since the drawers display panel will have the MTMS and not frame address displayed, a function is provided to display the frame address.

- If the AC power is not applied, then apply it now.
- The drawer may power up automatically.
- FRU replacement will generate a new temporary unit value in the expansion unit control panel. Use this new value to power down the expansion drawer without removing the power cord. See Powering off an expansion unit, then return here and continue with the next step.
- Select *function 07* on the control panel and press Enter.
- Select *sub function A6* to display the address of all units. The frame address is displayed on all units for 30 seconds.
- Note the frame address on the unit that you are working on for use in the next steps.
- Select *sub function A9* to set the ID of a drawer.
- Use the arrow keys to increment/decrement to the first two digits of the frame address noted above.
- Press Enter.
- Use the arrow keys to increment/decrement to the last two digits of the frame address noted above.
- Press Enter.
- Use the arrow keys to increment/decrement to a configuration ID for the type of unit you are working on:
  - 81 for 5074 and 5079 expansion units
  - 89 for 5088 and 0588 expansion units
  - 8A for 5094 and 5294 expansion units

   - 8B for 5095 and 0595 expansion units
   - 88 for 7311-D10, 7311-D11, and 5790 expansion units
   - 8C for 7311-D20 expansion units
- Press Enter (078x 00 will be displayed).
- Use the arrow keys to increment/decrement until 07** is shown.
- Press Enter to return the panel to 07.
- Disconnect all AC power to the unit, wait for the display panel to go off and then reconnect the AC power.
- The drawer will automatically power on.
- Continue with the next step to update the MTMS value using the ASMI. If you do not have access to the ASMI, then return to the procedure that sent you here.
- On the ASMI Welcome pane, specify your user ID and password, and click Log In.
- In the navigation area, expand System Configuration and click Configure I/O Enclosures.
- Select the unit identified by the new value in the panel of the unit you are working on. In most cases it will appear as *TMPx.yyy.yyyyyyy* where x is a hex digit, and the *y's* are any value.
- Select change settings.
- Enter the type-model from the label on the I/O unit
- Enter the serial number (also called sequence number) from the label on the I/O unit.
- Click Save Setting to complete the operation.

> **Note:** Do not use the browser back button or the values will not be saved. Verify the correct value is now displayed in the panel of the unit you are working on. Log off and close the ASMI. Then return to the procedure that sent you here.

3

POWER5 virtualization capabilities

Virtualization is a critical component in the on demand operating environment, and the system technologies implemented in the POWER5 processor-based IBM eServer p5 servers provide a significant advancement in the implementation of functions required for operating in this environment. IBM virtualization innovations on the p5-590 and p5-595 provide industry-unique utilization capabilities for a more responsive, flexible, and simplified infrastructure.

Advanced POWER Virtualization is a no-charge feature (FC 7992) on the p5-590 and p5-595 systems. On other p5 systems, it is a priced feature.

The Advanced POWER Virtualization feature provides the foundation of POWER5 virtualization technology. In this chapter we introduce the POWER5 Virtualization Engine and associated features, including Micro-Partitioning, shared processor pooling, virtual I/O (disk and LAN), and the Partition Load Manager for AIX 5L logical partitions.

3.1 p5 virtualization features

IBM Virtualization Engine, as shown in Figure 3-1, is comprised of a suite of system services and technologies that form key elements of IBMs on demand computing model. It treats resources of individual servers, storage, and networking products to function as a single pool, allowing access and management of resources across an organization more efficiently. Virtualization is a critical component in the on demand operating environment, and the system technologies implemented in the POWER5 processor-based IBM eserver p5 servers provide a significant advancement in the enablement of functions required for operating in this environment.

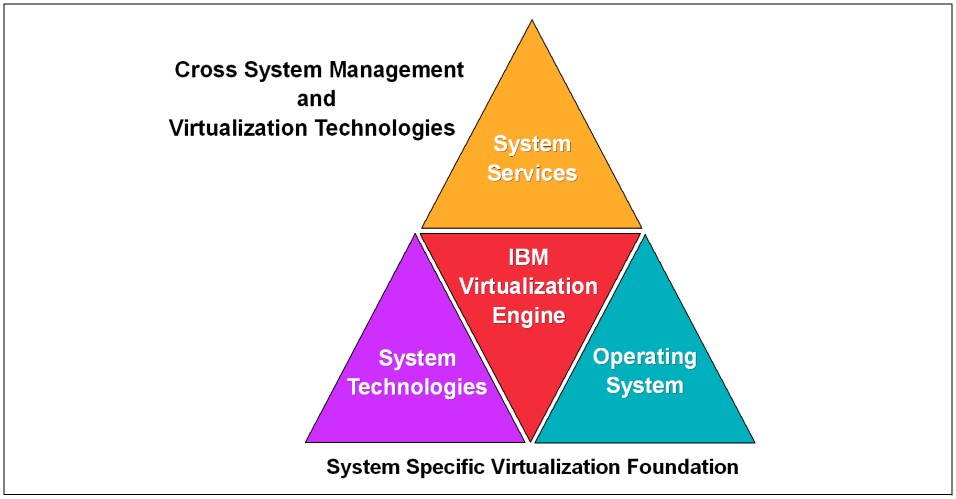

Cross System Management
and
Virtualization Technologies
System Services
IBM Virtualization Engine
System Technologies
Operating System
System Specific Virtualization Foundation

Figure 3-1 IBM Virtualization Engine components

The following sections explain the virtualization engine system technologies that are integrated into eserver p5 system hardware and operating systems, including:

Micro-Partitioning Enables you to allocate less than a full physical processor to a logical partition allowing increased overall resource utilization

Virtual Ethernet Provides network virtualization capabilities that allows communications between integrated servers

Virtual I/O Provides the ability to dedicate I/O adapters and devices to a virtual server, allowing the on demand allocation and management of I/O devices

POWER Hypervisor Supports partitioning and dynamic resource movement across multiple operating system environments

Figure 3-2 shows how several of these technologies combine to provide you the flexibility to help meet your computing requirements.

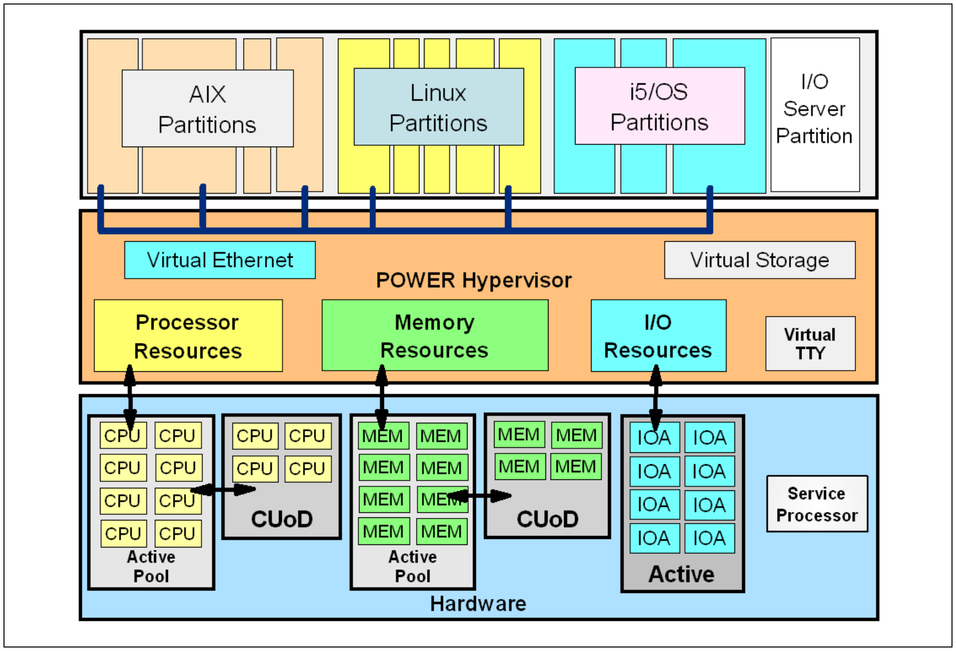


Figure 3-2 Virtualization technologies implemented on POWER5 servers

The following reference is recommended for the reader that is looking for more introductory material on IBM concepts on virtualization: *Advanced POWER Virtualization on IBM @server p5 Servers Introduction and Basic Configuration*, SG24-7940.

3.2 Micro-Partitioning

Micro-Partitioning is an advanced virtualization feature of POWER5 systems with AIX 5L Version 5.3 and Linux (SUSE LINUX Enterprise Server 9 for POWER systems and Red Hat Enterprise Linux AS 3 for POWER) that allows multiple partitions to share the processing power of a set of physical processors. A partition can be assigned as little as 1/10th of a physical processor's resource. The POWER Hypervisor controls the dispatching of the physical processors to each of the shared processor partitions. In most cases, a shared processor pool containing multiple physical processors is shared among the partitions. Shared processor partitions (partitions using Micro-Partitioning technology) still need dedicated memory, but the partition's I/O requirements can be supported though Virtual Ethernet adapter and Virtual SCSI server. Virtual Ethernet and Virtual

SCSI are briefly explained in Section 3.3, “Virtual Ethernet” on page 64 and Section 3.6, “Virtual SCSI” on page 80. Micro-Partitioning requires the Advanced POWER Virtualization capabilities.

3.2.1 Shared processor partitions

The virtualization of processors enables the creation of a partitioning model which is fundamentally different from the POWER4 systems where whole processors are assigned to partitions and are owned by them. In the new 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 microcode. From an operating system perspective, a virtual processor is indistinguishable from a physical processor. The key benefit of implementing partitioning in the hardware is to allow any operating system to run on POWER5 technology with little or no changes. Optionally, for optimal performance, the operating system can be enhanced to exploit shared processor pools more in-depth. For instance, by voluntarily relinquishing CPU cycles to the hardware when they are not needed. AIX 5L Version 5.3 is the first version of AIX 5L that includes such enhancements.

Micro-Partitioning allows for multiple partitions to share one physical processor. Partitions using Micro-Partitioning technology are referred to as shared processor partitions.

A partition may be defined with a processor capacity as small as 10 processor units. This represents 1/10 of a physical 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 control of the POWER Hypervisor.

Figure 3-3 on page 59 shows the POWER5 partitioning concept:

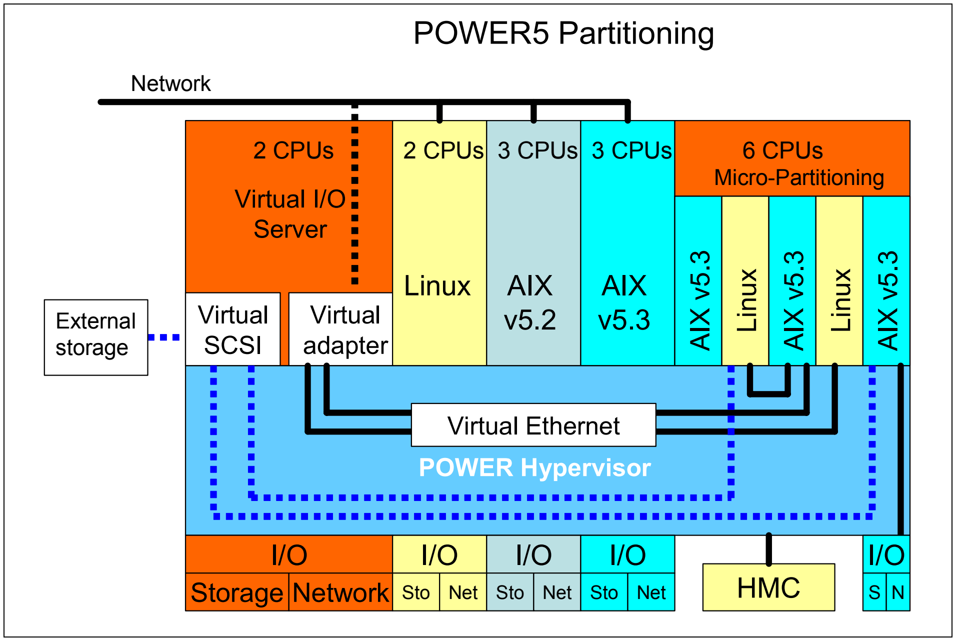


Figure 3-3 POWER5 partitioning concept

Micro-Partitioning is supported across the entire POWER5 product line from the entry to the high-end systems. Table 3-1 provides the maximum number of logical partitions and shared processor partition the different models.

Table 3-1 Micro-Partitioning overview on p5 systems

p5 servers	**Model 510**	**Model 520**	**Model 550**	**Model 570**	**Model 590**	**Model 595**
Processors	2	2	4	16	32	64
Dedicated processor partitions	2	2	4	16	32	64
Shared processor partitions	20	20	40	160	254	254

It is important to point out that the maximums stated are supported by the hardware, but the practical limits based on production workload demands may be significantly lower.

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.

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, where other partitions require greater power.

From 0.1 to 1 processor units can be used by one virtual processor according to different setting. By default, the number of virtual processor will be automatically set to the minimum number of virtual processors needed to satisfy the assigned number of processor unit. The default settings maintains 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 partitions profile to change the default configuration and to assign more virtual processors.

At the time of publication, the maximum number of virtual processors per partition is 64.

Dedicated processors

Dedicated processors are whole processors that are assigned to a single partition. If you choose to assign dedicated processors to a logical partition, you must assign at least one processor to that partition.

You can not mix shared processors and dedicated processors in one partition.

By default, a powered-off logical partition using dedicated processors will have its processors available to the shared processing pool. When the processors are in the shared processing pool, an uncapped partition that needs more processing power can use the idle processing resources. However, when you power on the dedicated partition while the uncapped partition is using the processors, the activated partition will regain all of its processing resources. If you want to prevent dedicated processors from being used in the shared processing pool, you can disable this function using the logical partition profile properties panels on the Hardware Management Console.

Note: You cannot disable the *allow idle processor to be shared* function when you create a partition. You need to open the properties for the created partition and change it on the processor tab.

3.2.2 Types of shared processor partitions

Shared processor partitions can be of two different types, depending on the capacity they have of using idle processing resources available on the system. If a processor donates unused cycles back to the shared pool, or if the system has idle capacity (because there is not enough workload running), the extra cycles may be used by some partitions, depending on their type and configuration.

Capped partitions

A capped partition is defined with a hard maximum limit of processing capacity. That means that it cannot go over its defined maximum capacity in any situation, unless you change the configuration for that partition (either by modifying the partition profile or by executing a DLPAR operation). Even if the system is idle, the capped partition may reach a processor utilization of 100%.

Figure 3-4 on page 61 shows an example where a shared processor partition is capped at an entitlement of 9.5 (up to the equivalent of 9.5 physical processors). In some moments the processor usage goes up to 100 percent, and while the machine presents extra capacity not being used, by design the capped partition cannot use it.

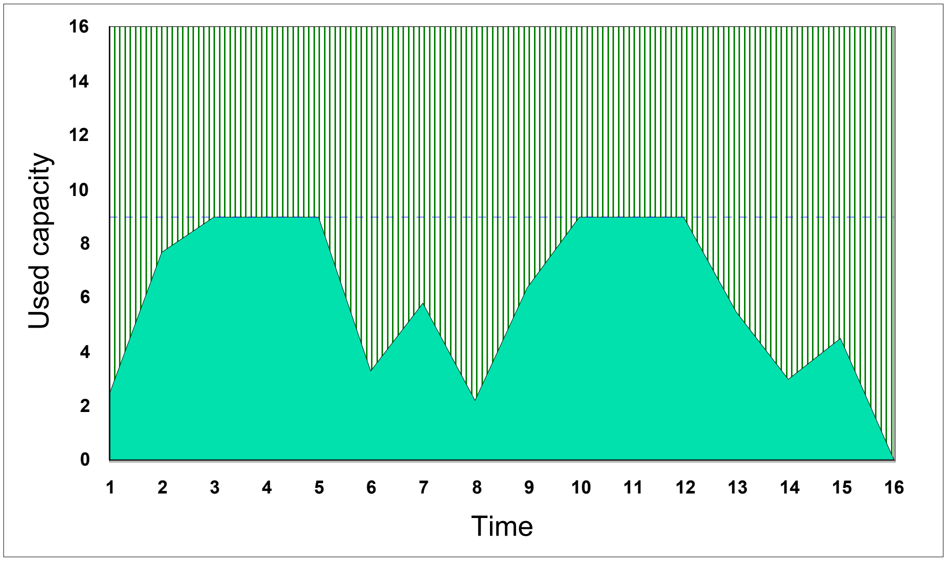


Figure 3-4 Capped shared processor partitions

Uncapped partitions

An uncapped partition has the same definition of a capped partition, except that the maximum limit of processing capacity limit is a soft limit. That means that an uncapped partition may eventually receive more processor cycles than its entitled capacity.

In the case it is using 100 percent of the entitled capacity, and there are idle processors in the shared processor pool, the POWER Hypervisor has the ability to dispatch virtual processors from the uncapped partitions to use the extra capacity.

In the example we used for the capped partition, if we change the partition from capped to uncapped, a possible chart for the capacity utilization is the one shown in Figure 3-5 on page 62. It still has the equivalent of 9.5 physical processors as its entitlement, but it an use more resources if required.

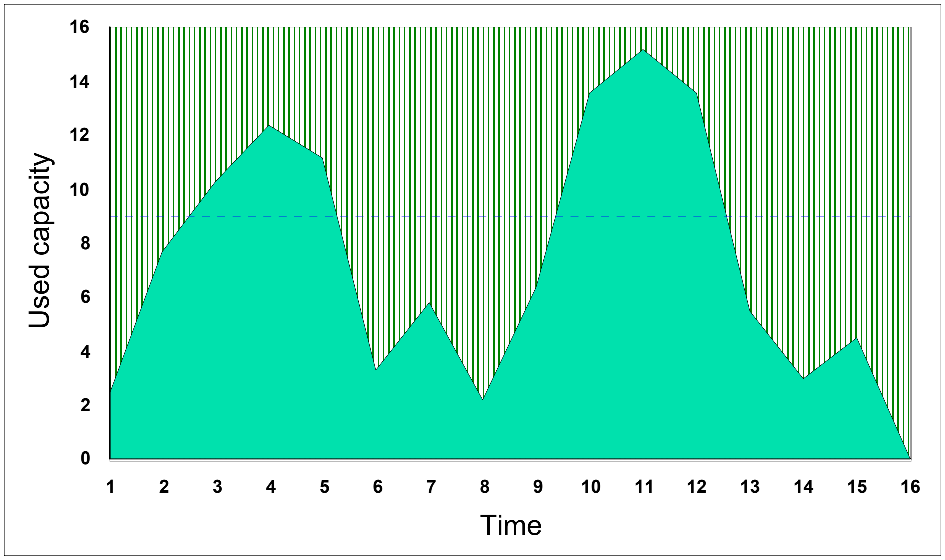


Figure 3-5 Uncapped shared processor partitions

The number of virtual processors on an uncapped partition defines the largest capacity it can use from the shared pool. By making the number of virtual processors too small, you may limit the processing capacity of an uncapped partition. A logical partition in the shared processing pool will have at least as many virtual processors as its assigned processing capacity. If you have a partition with 0.50 processing units and 1 virtual processor, the partition can not 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 could use an additional 1.50 processing units.

3.2.3 Typical usage of Micro-Partitioning technology

With fractional processor allocations, more partitions can be created on a given platform which enable clients to maximize the number of workloads that can be supported simultaneously. Micro-Partitioning enables both optimized use of processing capacity while preserving the isolation between applications provided by different operating system images.

There are several scenarios where the usage of Micro-Partitioning can bring advantages such as optimal resource utilization, rapid deployment of new servers and application isolation:

- **Server consolidation**

  Consolidating several small systems onto a large and robust server brings advantages in management and performance, usually together with reduced total cost of ownership (TCO). A Micro-Partitioning system enables the consolidation from small to large systems without the burden of dedicating very powerful processors to a small partition. You can divide the processor between several partitions with the adequate processing capacity for each one.

- **Server provisioning**

  With Micro-Partitioning and virtual I/O, a new partition can be deployed rapidly, to accommodate unplanned demands, or to be used as a test environment.

- **Virtual server farms**

  n environments where applications scale with the addition of new servers, the ability to create several partitions sharing processing resources is very useful and contributes for a better use of processing resources by the applications deployed on the server farm.

3.2.4 Limitations and considerations

The following limitations should be considered when implementing shared processor partitions:

- The limitation for a shared processor partition is 0.1 processing units of a physical processor. So the number of shared processor partitions you can create for a system depends mostly on the number of processors of a system.
- The maximum number of partitions is 254.
- In a partition, there is a maximum number of 64 virtual processors.
- A mix of dedicated and shared processors within the same partition is not supported.

- If you dynamically remove a virtual processor you can not specify a particular virtual CPU to be removed. The operating system will choose the virtual CPU to be removed.
- Shared processors may render AIX affinity management useless. AIX will continue to utilize affinity domain information as provided by firmware to build associations of virtual processors to memory, and will continue to show preference to redispatching a thread to the virtual CPU that it last ran on.

Operating systems and applications running in shared partitions need not be aware that they are sharing processors. However, overall system performance can be significantly improved by minor operating system changes. AIX 5L Version 5.3 provides support for optimizing overall system performance of shared processor partitions.

In a shared partition, there is not a fixed relationship between the virtual processor and the physical processor. The POWER Hypervisor will try to use a physical processor with the same memory affinity as the virtual processor, but it is not guaranteed. Virtual processors have the concept of a home physical processor. If it can't find a physical processor with the same memory affinity, then it gradually broadens its search to include processors with weaker memory affinity, until it finds one that it can use. As a consequence, memory affinity is expected to be weaker in shared processor partitions.

Workload variability is also expected to be increased in shared partitions, because there are latencies associated with the scheduling of virtual processors and interrupts. simultaneous multi-threading may also increase variability, since it adds another level of resource sharing, which could lead to a situation where one thread interferes with the forward progress of its sibling.

If an application is cache sensitive or cannot tolerate variability, then the dedicated partition with simultaneous multi-threading disabled is recommended. In dedicated partitions, the entire processor is assigned to a partition. Processors are not shared with other partitions, and they are not scheduled by the POWER Hypervisor. Dedicated partitions must be explicitly created by the system administrator using the HMC.

3.3 Virtual Ethernet

Virtual Ethernet enables inter-partition communication without the need for physical network adapters assigned to each partition. Virtual Ethernet allows the administrator to define in-memory point to point connections between partitions. These connections exhibit similar characteristics as physical high-bandwidth Ethernet connections and supports multiple protocols (IPv4, IPv6, ICMP). Virtual Ethernet requires a IBM @server p5 590 and p5 595 system with either AIX 5L

Version 5.3 or the appropriate level of Linux and a Hardware Management Console (HMC) to define the Virtual Ethernet devices. Virtual Ethernet does not require the Advanced POWER Virtualization feature.

Due to the number of partitions possible on many systems being greater than the number of I/O slots, Virtual Ethernet is a convenient and cost saving option to enable partitions within a single system to communicate with one another through a VLAN. The VLAN creates logical Ethernet connections between one or more partitions and is designed to help avoid a failed or malfunctioning operating system from being able to impact the communication between two functioning operating systems. The Virtual Ethernet connections may also be *bridged* to an external network to permit partitions without physical network adapters to communicate outside of the server.

The concepts of implementing Virtual Ethernet are categorized in the following sections:

- Section 3.3.1, “Virtual LAN” on page 65
- Section 3.3.2, “Virtual Ethernet connections” on page 69
- Section 3.3.3, “Dynamic partitioning for virtual Ethernet devices” on page 71
- Section 3.3.4, “Limitations and considerations” on page 71

3.3.1 Virtual LAN

This section will discuss the concepts of Virtual LAN (VLAN) technology with specific reference to its implementation within AIX.

Virtual LAN overview

Virtual LAN (VLAN) is a technology used for establishing virtual network segments on top of physical switch devices. If configured appropriately, a VLAN definition can straddle multiple switches. 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. In Figure 3-6, two VLANs (VLAN 1 and 2) are defined on three switches (Switch A, B, and C). Although nodes C-1 and C-2 are physically connected to the same switch C, traffic between two nodes can be blocked. To enable communication between VLAN 1 and 2, L3 routing or inter-VLAN bridging should be established between them; this is typically provided by an L3 device.

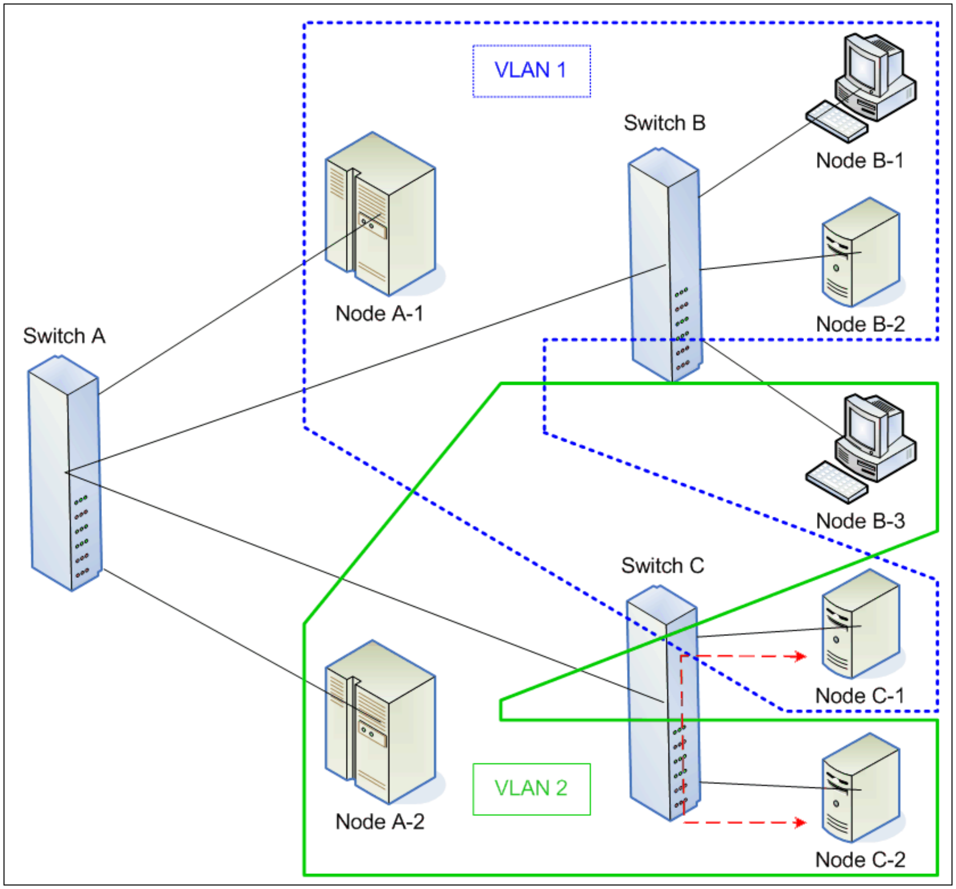


Figure 3-6 Example of a VLAN

The use of VLAN provides increased LAN security and flexible network deployment over traditional network devices.

AIX virtual LAN support

Some of the various technologies for implementing VLANs include:

- Port-based VLAN
- Layer 2 VLAN
- Policy-based VLAN
- IEEE 802.1Q VLAN

VLAN support in AIX is based on the IEEE 802.1Q VLAN implementation. The IEEE 802.1Q VLAN is achieved by adding a VLAN ID tag to an Ethernet frame, and the Ethernet switches restricting the frames to ports that are authorized to receive frames with that VLAN ID. Switches also restrict broadcasts to the logical network by ensuring that a broadcast packet is delivered to all ports which are configured to receive frames with the VLAN ID that the broadcast frame was tagged with.

A port on a VLAN capable switch has a default PVID (Port virtual LAN ID) that indicates the default VLAN the port belongs to. The switch adds the PVID tag to untagged packets that are received by that port. In addition to a PVID, a port may belong to additional VLANs and have those VLAN IDs assigned to it that indicates the additional VLANs the port belongs to.

A port will only accept untagged packets or packets with a VLAN ID (PVID or additional VIDs) tag of the VLANs the port belongs to. 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. Each VLAN logical devices constitutes an additional Ethernet adapter instance. These logical devices can be used to configure the same Ethernet IP interfaces as are used with physical Ethernet adapters.

VLAN communication by example

This section discusses how VLAN communication between partitions and with external networks works in more detail using the sample configuration in Figure 3-7. The configuration is using four client partitions (Partition 1 - 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 which bridges traffic to the external network. The Shared Ethernet Adapter will be introduced in more detail in Chapter 3.4 Shared Ethernet adapter.

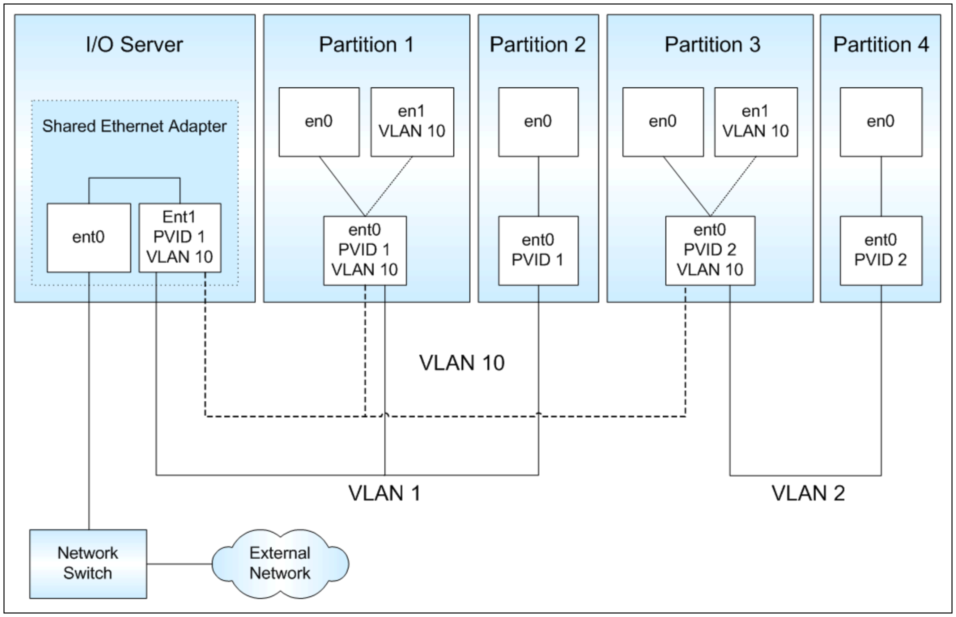


Figure 3-7 VLAN configuration

Interpartition communication

Partition 2 and Partition 4 are using the PVID (Port virtual LAN ID) only. This means that:

- Only packets for the VLAN specified as PVID are received
- Packets sent are added a VLAN tag 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 specific network interface (en1) create through **`smitty vlan`**. This means that:

- Packets sent through network interfaces en1 are added a tag for VLAN 10 by the network interface in AIX.
- 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 3-2 lists which client partition can communicate which each other through what network interfaces.

Table 3-2 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 which 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 network interface en0 using VLAN 1. As 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 network interface en1 and VLAN 10. This 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 3-3 lists this relationship.

Table 3-3 VLAN communication to external network

VLAN	Partition / Network interface
1	Partition 1 / en0 Partition 2 / en0
10	Partition 1 / en1 Partition 3 / en1

3.3.2 Virtual Ethernet connections

Virtual Ethernet connections supported in POWER5 systems use VLAN technology to ensure that the partitions can only access data directed to them.

The POWER Hypervisor provides a Virtual Ethernet switch function based on the IEEE 802.1Q VLAN standard that allows partition communication within the same server. The connections are based on an implementation internal to the hypervisor that moves data between partitions. This section will describe the various elements of a Virtual Ethernet and implications relevant to different types of workloads. Figure 3-8 is an example of an inter-partition VLAN.

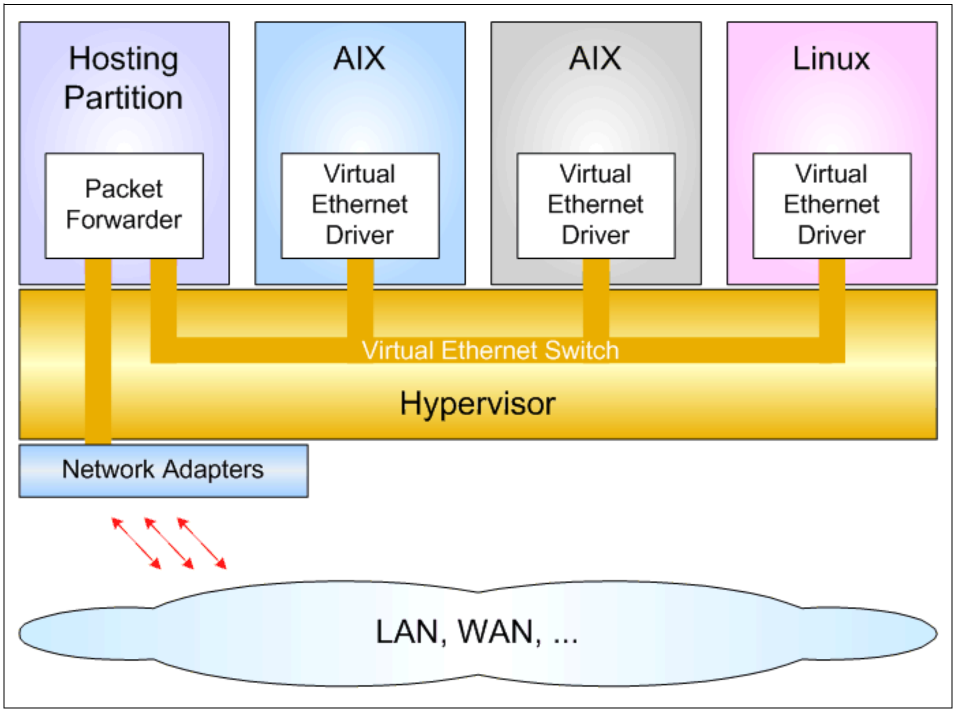


Figure 3-8 Logical view of an inter-partition VLAN

Virtual Ethernet adapter concepts

Partitions that communicate through a Virtual Ethernet channel will need to have an additional in-memory channel. This requires the creation of an in-memory channel between partitions on the HMC. The kernel creates a virtual device for each memory channel indicated by the firmware. The AIX configuration manager creates the device special files. A unique Media Access Control (MAC) address is also generated when the Virtual Ethernet device is created. A *prefix* value can be assigned for the system so that the generated MAC addresses in a system consists of a common system prefix, plus an algorithmically-generated unique part per adapter.

The Virtual Ethernet can also be used as a bootable device to allow such tasks as operating system installations to be performed using Network Installation Management (NIM).

Performance considerations

The transmission speed of Virtual Ethernet adapters is in the range of 1-3 Gigabits per second, depending on the transmission (MTU) size. A partition can support up to 256 Virtual Ethernet adapters with each Virtual Ethernet capable to be associated with up to 21 VLANs (20 VID and 1 PVID).

The Virtual Ethernet connections generally take up more processor time than a local adapter to move a packet (DMA versus copy). For shared processor partitions, performance will be gated by the partition definitions (for example entitled capacity and number of processors). Small partitions communicating with each other will experience more packet latency due to partition context switching. In general, high bandwidth applications *should not* be deployed in small shared processor partitions. For dedicated partitions, throughput *should be* comparable to a 1 Gigabit Ethernet for small packets providing much better performance than 1 Gigabit Ethernet for large packets. For large packets, the Virtual Ethernet communication is copy bandwidth limited.

For more detailed information relating to Virtual Ethernet performance considerations refer to the following publication:

- *Advanced POWER Virtualization on IBM eServer p5 Servers Architecture and Performance Considerations*, SG24-5768.

3.3.3 Dynamic partitioning for virtual Ethernet devices

Virtual Ethernet resources can be assigned and removed dynamically. On the HMC, Virtual Ethernet target and server adapters can be assigned and removed from a partition using dynamic logical partitioning. The mapping between physical and virtual resources on the Virtual I/O Server can also be done dynamically.

3.3.4 Limitations and considerations

The following are limitations that must be considered when implementing an Virtual Ethernet:

- A maximum of up to 256 Virtual Ethernet adapters are permitted per partition.
- Virtual Ethernet can be used in both shared and dedicated processor partitions provided the partition is running AIX 5L Version 5.3 or Linux with the 2.6 kernel or a kernel that supports virtualization.

- A mixture of Virtual Ethernet connections, real network adapters, or both are permitted within a partition.
- Virtual Ethernet can only connect partitions within a single system.
- Virtual Ethernet connections from AIX or Linux partitions to an i5/OS partition may work, however at time of writing, these capabilities were unsupported.
- Virtual Ethernet uses the system processors for all communication functions instead of off loading that load to processors on network adapter cards. As a result there is an increase in system processor load generated by the use of Virtual Ethernet.

3.4 Shared Ethernet Adapter

A Shared Ethernet Adapter can be used to connect a physical Ethernet to the Virtual Ethernet. It also provides the possibility for several client partitions to share one physical adapter.

The following sections discuss the various aspects of Shared Ethernet Adapters such as:

- Section 3.4.1, “Connecting a virtual Ethernet to external networks” on page 72
- Section 3.4.2, “Using Link Aggregation (EtherChannel) to external networks” on page 75
- Section 3.4.3, “Limitations and considerations” on page 77

3.4.1 Connecting a virtual Ethernet to external networks

There are two ways you can connect the Virtual Ethernet that enables the communication between logical partitions on the same server to an external network.

Routing

By enabling the AIX routing capabilities (ipforwarding network option) one partition with a physical Ethernet adapter connected to an external network can act as router. Figure 3-9 shows a sample configuration. In this type of configuration the partition that routes the traffic to the external work does not necessarily have to be the Virtual I/O Server as in the example below. It could be any partition with a connection to the outside world. The client partitions would have their default route set to the partition which routes traffic to the external network.

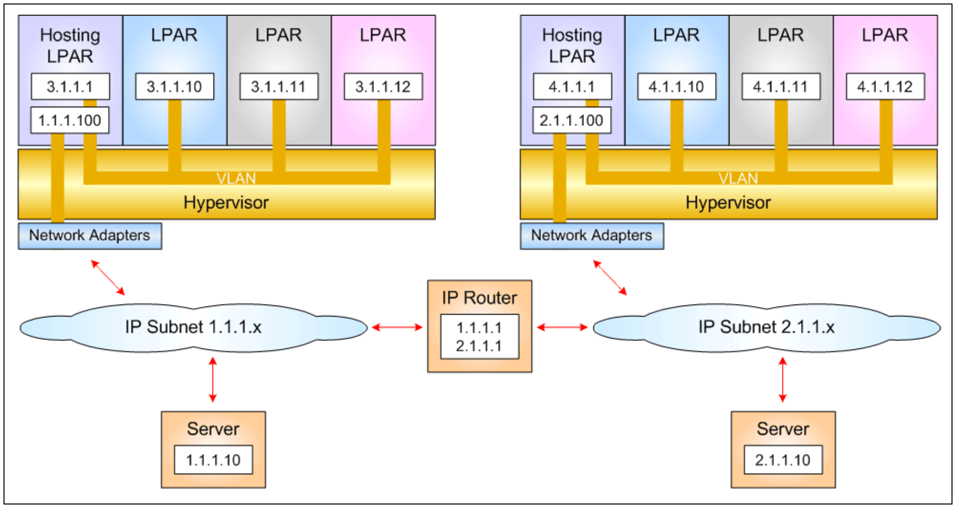


Figure 3-9 Connection to external network using AIX routing

Shared Ethernet Adapter

Using a Shared Ethernet Adapter (SEA) you can connect internal and external VLANs using one physical adapter. The Shared Ethernet Adapter hosted in the Virtual I/O Server acts as a layer 2 switch between the internal and external network.

Shared Ethernet Adapter is a new service that acts as a layer 2 network bridge to securely transport network traffic from a virtual Ethernet to a real network adapter. The Shared Ethernet Adapter service runs in the Virtual I/O Server. It cannot be run in a general purpose AIX partition.

Shared Ethernet Adapter requires the POWER Hypervisor and Advanced POWER Virtualization features of POWER5 systems and therefore cannot be used on POWER4 systems. It also cannot be used with AIX 5L Version 5.2 because the device drivers for virtual Ethernet are only available for AIX 5L Version 5.3 and Linux. Thus there is no way to connect an AIX 5L Version 5.2 system to a Shared Ethernet Adapter.

The Shared Ethernet Adapter allows partitions to communicate outside the system without having to dedicate a physical I/O slot and a physical network adapter to a client partition. The Shared Ethernet Adapter has the following characteristics:

- Virtual Ethernet MAC addresses are visible to outside systems
- Broadcast/multicast is supported
- ARP and NDP can work across a shared Ethernet

In order to bridge network traffic between the Virtual Ethernet and external networks the Virtual I/O Server has to be configured with at least one physical Ethernet adapter. One Shared Ethernet Adapter can be shared by multiple VLANs and multiple subnets can connect using a single adapter on the Virtual I/O Server. Figure 3-10 on page 74 shows a configuration example. A Shared Ethernet Adapter can include up to 16 Virtual Ethernet adapters that share the physical access.

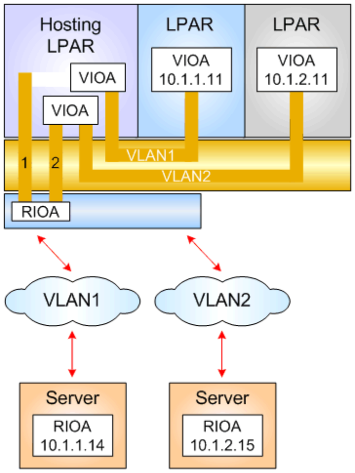


Figure 3-10 Shared Ethernet Adapter configuration

In the LPAR profile for the VIO Server partition, the Virtual Ethernet adapter which will be associated with the (physical) Shared Ethernet Adapter must have the trunk flag set. Once an Ethernet frame is sent from the Virtual Ethernet adapter on a client partition to the POWER Hypervisor, the POWER Hypervisor searches for the destination MAC address within the VLAN. If no such MAC address exists within the VLAN, it forwards the frame to the trunk Virtual Ethernet adapter that is defined on the same VLAN. The trunk virtual Ethernet adapter enables a layer 2 bridge to a physical adapter.

The shared Ethernet directs packets based on the VLAN ID tags. It learns this information based on observing the packets originating from the virtual adapters. One of the virtual adapters in the Shared Ethernet adapter is designated as the default PVID adapter. Ethernet frames without any VLAN ID tags are directed to this adapter and assigned the default PVID.

When the shared Ethernet receives IP (or IPv6) packets that are larger than the MTU of the adapter that the packet is forwarded through, either IP fragmentation is performed and the fragments forwarded or an ICMP packet too big message is returned to the source when the packet cannot be fragmented.

Theoretically, one adapter can act as the only contact with external networks for all client partitions. Depending on the number of client partitions and the network load they produce performance can become a critical issue. Because the Shared Ethernet Adapter is dependant on virtual I/O, it consumes processor time for all communications. A significant amount of CPU load can be generated by the use of Virtual Ethernet and Shared Ethernet Adapter.

There are several different ways to configure physical and virtual Ethernet adapters into Shared Ethernet Adapters to maximize throughput.

- Using Link Aggregation (EtherChannel), several physical network adapter can be aggregated. See 3.4.2, “Using Link Aggregation (EtherChannel) to external networks” on page 75 for more details.
- Using several Shared Ethernet Adapters provides more queues and more performance. An example for this configuration is shown in Figure 3-11.

Other aspects which have to be taken into consideration are availability and the possibility to connect to different networks.

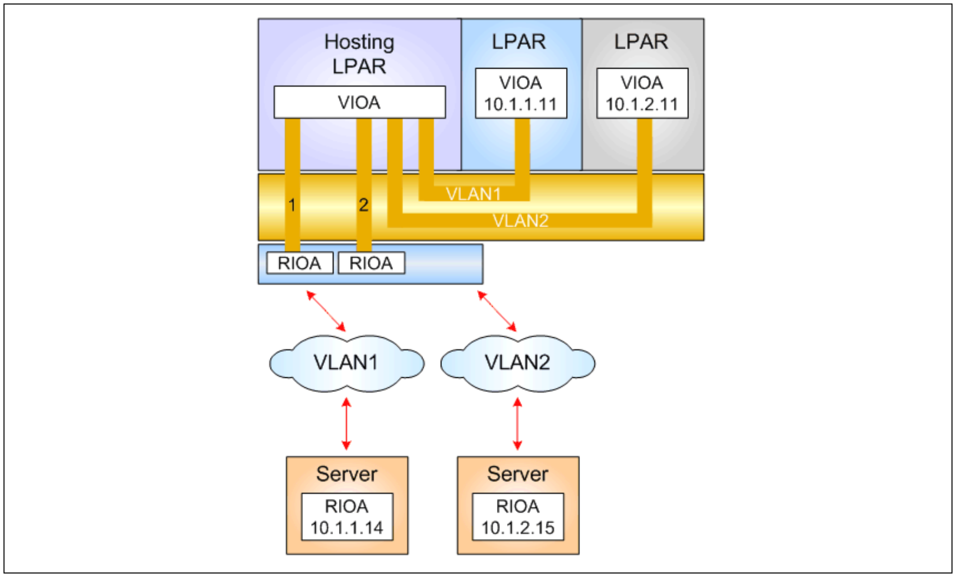


Figure 3-11 Multiple Shared Ethernet Adapter configuration

3.4.2 Using Link Aggregation (EtherChannel) to external networks

Link aggregation is network port aggregation technology that allows several Ethernet adapters to be aggregated together to form a single pseudo Ethernet device. This technology can be used to overcome the bandwidth limitation of a single network adapter and avoid bottlenecks when sharing one network adapter amongst many client partitions.

For example, ent0 and ent1 can be aggregated to ent3. Interface en3 would then be configured with an IP address. The system considers these aggregated adapters as one adapter. Therefore, IP is configured as on any other Ethernet adapter. In addition, all adapters in the link aggregation are given the same hardware (MAC) address, so they are treated by remote systems as if they were one adapter. The main benefit of link aggregation is that they have the network bandwidth of all of their adapters in a single network presence. If an adapter fails, the packets are automatically sent on the next available adapter without disruption to existing user connections. The adapter is automatically returned to service on the link aggregation when it recovers.

You can use EtherChannel (EC) or IEEE 802.3ad Link Aggregation (LA) to aggregate network adapters. While EC is an AIX specific implementation of adapter aggregation, LA follows the IEEE 802.3ad standard. Table 3-4 shows the main differences between EC and LA.

Table 3-4 Main differences between EC and LA aggregation

EtherChannel	IEEE 802.3ad link aggregation
Requires switch configuration	Little, if any, configuration of switch required to form aggregation. Some initial setup of the switch may be required.
Supports different packet distribution modes	Supports only standard distribution mode

The main benefit of using LA is, that if the switch supports the *Link Aggregation Control Protocol* (LACP) no special configuration of the switch ports is required. The benefit of EC is the support of different packet distribution modes. This means it is possible to influence the load balancing of the aggregated adapters. In the remainder of this document, we will use Link Aggregation where possible since that is considered a more universally understood term.

Note: Only outgoing packets are subject to the following discussion, incoming packets are distributed by the Ethernet switch.

Standard distribution mode selects the adapter for the outgoing packets by algorithm. The adapter selection algorithm uses the last byte of the destination IP address (for TCP/IP traffic) or MAC address (for ARP and other non-IP traffic). Therefore all packets to a specific IP-address will always go through the same adapter. There are other adapter selection algorithms based on source, destination, or a combination of source and destination ports available. EC provides one further distribution mode called *round robin.* This mode will rotate through the adapters, giving each adapter one packet before repeating. The packets may be sent out in a slightly different order than they were given to the EC. It will make the best use of its bandwidth, but consider that it also introduces the potential for out-of-order packets at the receiving system. This risk is particularly high when there are few, long-lived, streaming TCP connections. When there are many such connections between a host pair, packets from different connections could be intermingled, thereby decreasing the chance of packets for the same connection arriving out-of-order.

To avoid the loss of network connectivity by switch failure, EC and LA can provide a backup adapter. The backup adapter should be connected to different a switch than the adapter of the aggregation. Now in case of switch failure the traffic can be moved with no disruption of user connections to the backup adapter.

Figure 3-12 shows the aggregation of three plus one adapters to a single pseudo Ethernet device including a backup feature.

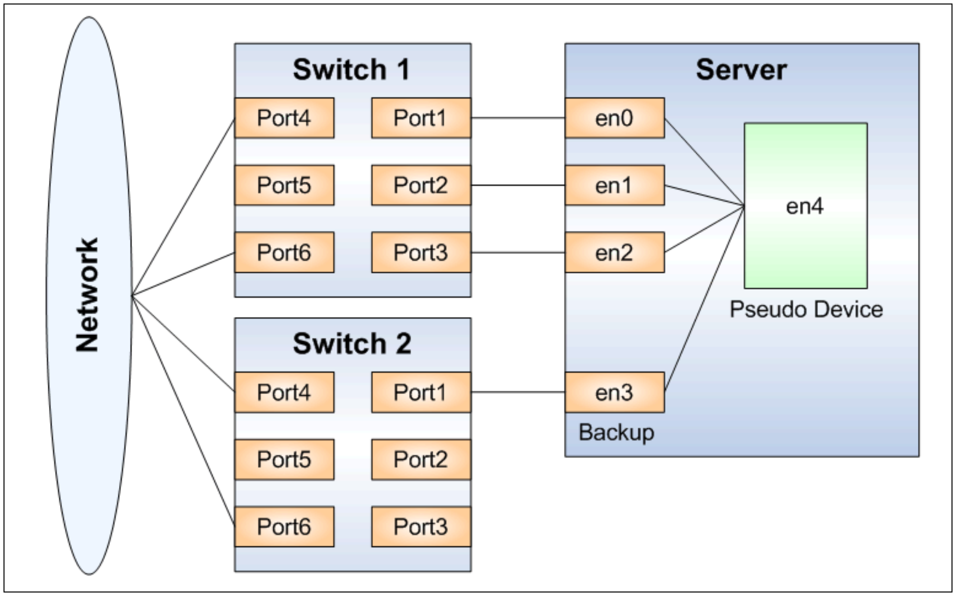


Figure 3-12 Link Aggregation (EtherChannel) pseudo device

3.4.3 Limitations and considerations

The following limitations you must consider when implementing Shared Ethernet Adapters in the Virtual I/O Server:

- Because Shared Ethernet Adapter depends on virtual Ethernet which uses the system processors for all communications functions, a significant amount of system processor load can be generated by the use of virtual Ethernet and Shared Ethernet Adapter.
- One of the virtual adapters in the Shared Ethernet Adapter on the Virtual I/O Server must be defined as default adapter with a default PVID. This virtual adapter is designated as the PVID adapter and Ethernet frames without any VLAN ID tags are assigned the default PVID and directed to this adapter.
- Up to 16 virtual Ethernet adapters with 21 VLANs (20 VID and 1 PVID) on each can be shared on a single physical network adapter. There is no limit on the number of partitions that can attach to a VLAN. So the theoretical limit is very high. In practice, the amount of network traffic will limit the number of clients that can be served through a single adapter.

3.5 Virtual I/O Server

The Advanced POWER Virtualization feature is included in the default configuration of the p5-590 and p5-595 systems. This feature includes Micro-Partitioning, shared processor pooling, virtual I/O, and Partition Load Manager for AIX 5L logical partitions.

The IBM Virtual I/O Server (VIOS) 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 such as Ethernet adapters, SCSI disks, or Fibre Channel disks. Dedicated I/O resources are assigned to the VIOS, which allocates and manages I/O devices across multiple client partitions. The shared resource appears as a *virtual I/O device* or a virtual adapter to the client partition, as reflected in Figure 3-13 on page 79.

The client partition does not communicate directly with the underlying physical I/O adapters. Rather, when the client partition uses its virtual I/O device, it is sending operation requests to the Virtual I/O Server. The POWER Hypervisor provides a secure communication channel between the client partition and the VIOS, which transports the operation requests and responses between the client partition and the VIOS. When the VIOS receives an operation request for a virtual device from a client partition, that request is processed and mapped to an appropriate operation on a physical I/O device within the VIOS. This ability to provide virtual I/O resources enables the deployment of many partitions on a single server while reducing the number of physical adapters and devices required in the system configuration.

Figure 3-13 on page 79 depicts the relationship between the physical adapters in the VIOS and the virtual I/O adapters in the client partitions. The figure shows the full suite of POWER5 virtualization technologies, with additional detail regarding the Virtual I/O Server. In this drawing, the VIOS owns three physical I/O adapters; three client partitions own corresponding virtual I/O adapters. The thick solid lines running through the POWER Hypervisor represent the secure communication channels between the virtual I/O device in the client partition and the physical I/O device in the VIOS.

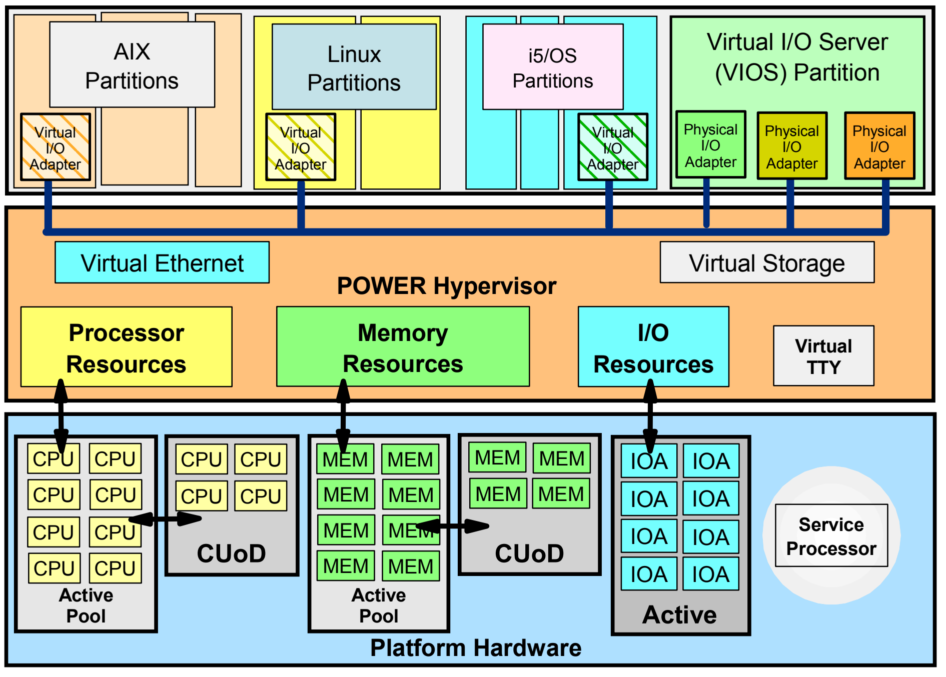


Figure 3-13 IBM p5-590 and p5-595 Virtualization Technologies

Virtual I/O Servers are the only partitions that can have virtual SCSI server adapters assigned to them. Any partition can have Virtual Ethernet adapters, and a VIOS is not necessarily required in order to use virtual Ethernet. However, the VIOS does provide a Shared Ethernet Adapter (SEA) service, which can bridge the internal virtual network traffic though a physical Ethernet adapter and out onto an external physical network. Virtual SCSI and Virtual Ethernet are briefly explained in Section 3.6, “Virtual SCSI” on page 80 and Section 3.3, “Virtual Ethernet” on page 64. Shared Ethernet functionality is described in Section 3.4, “Shared Ethernet Adapter” on page 72.

Although the VIOS is a component of the standard Advanced POWER Virtualization feature, which includes Shared Ethernet Adapter (SEA) service, the VIOS must be installed and configured by the client using the hardware management console (HMC). The VIOS is created using the HMC and is not intended for running applications or for general user logins. It is strictly designed to provide virtual I/O resources to logical client partitions.

3.6 Virtual SCSI

When each partition typically requires one I/O slot for disk attachment and another one for network attachment, this puts a constraint on the number of partitions. To overcome these physical limitations, I/O resources have to be shared. Virtual SCSI provides the means to do this for SCSI storage devices.

Virtual I/O allows the POWER5 to support more partitions than it has slots for I/O devices by enabling the sharing of I/O adapters amongst partitions. Virtual SCSI (VSCSI) will enable a partition to access block-level storage that is not a physical resource of that partition. The VSCSI design is that the virtual storage be backed by a logical volume on a portion of a disk rather than an entire physical disk. These logical volumes appear to be the SCSI disks on the client partition, which gives the system administrator maximum flexibility in configuring partitions.

Furthermore virtual I/O allows 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, if there is no native support for EMC storage devices on Linux, by running Linux in logical partition of a POWER5 server with virtual I/O make it possible.

Virtual I/O will provide a high performance I/O mechanism by minimizing the number of times data is copied within the memory of the physical system. The virtual I/O model described herein allows for either zero copy, if data is being retrieved from a physical device and DMAed directly to the memory of the partition using virtual I/O using the redirected DMA, or single copy of the data is first moved to the memory space of the I/O server before being DMAed to the I/O client.

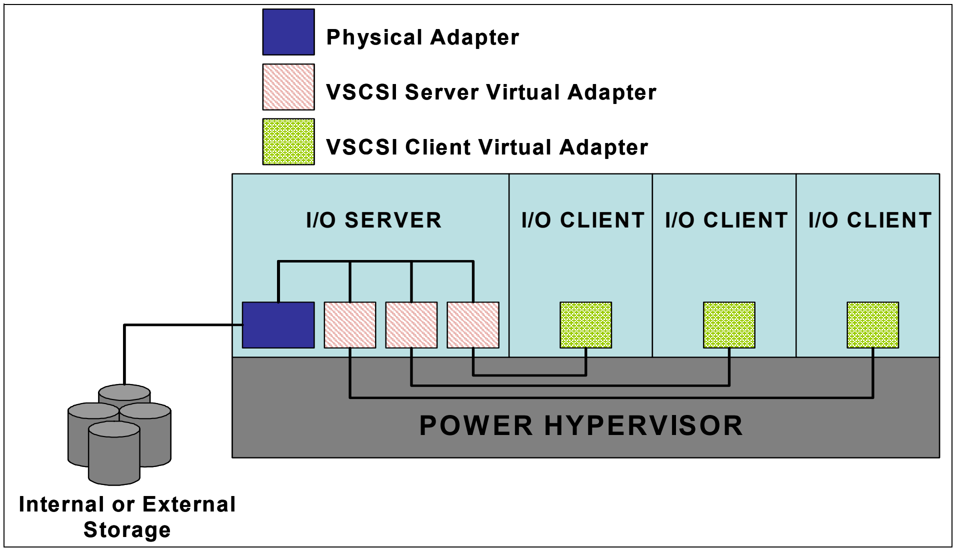

Physical Adapter
VSCSI Server Virtual Adapter
VSCSI Client Virtual Adapter
I/O SERVER
I/O CLIENT
I/O CLIENT
I/O CLIENT
POWER HYPERVISOR
Internal or External Storage

Figure 3-14 AIX 5L Virtual I/O Server and Client Partitions

3.6.1 Limitations and considerations

The following areas should be considered when implementing Virtual SCSI:

- At the time of writing virtual SCSI supports Fibre Channel, parallel SCSI, and SCSI RAID devices. Other protocols such as SSA or tape and CD-ROM devices are not supported.
- Virtual SCSI itself does not have any limitations in terms of number of supported devices or adapters. However, the Virtual I/O Server supports a maximum of 65535 virtual I/O slots. A maximum of 256 virtual I/O slots can be assigned to a single partition.

  Every I/O slot needs some resources to be instantiated. Therefore, the size of the Virtual I/O Server puts a limit to the number of virtual adapters that can be configured.
- The SCSI protocol defines mandatory and optional commands. While virtual SCSI supports all the mandatory commands, not all optional commands are supported. You can find a complete list of the supported SCSI commands in following publication: *Advanced POWER Virtualization on IBM p5 ⓔserver Introduction and Basic Configuration,* SG24-7940.
- There are performance implications when using Virtual SCSI devices. It is important to understand that associated with POWER Hypervisor calls, virtual SCSI will use additional CPU cycles when processing I/O requests. When putting heavy I/O load on virtual SCSI devices, this means you will use more

CPU cycles. Provided that there is sufficient CPU processing capacity available the performance of virtual SCSI should be comparable to dedicated I/O devices.

4

Capacity on Demand

The p5-590 and p5-595 systems can be shipped with inactive processor and memory resources, which may be activated at a future point in time without affecting normal machine operation. This ability is called *Capacity on Demand* (CoD) and provides flexibility and improved granularity in processor and memory upgrades.

This chapter includes the following topics:

- Section 4.1, “Capacity on Demand overview” on page 84
- Section 4.2, “What’s new in Capacity on Demand?” on page 84
- Section 4.3, “Preparing for Capacity on Demand” on page 85
- Section 4.4, “Types of Capacity on Demand” on page 87
- Section 4.5, “Capacity BackUp” on page 88
- Section 4.6, “Capacity on Demand activation procedure” on page 89
- Section 4.7, “Using Capacity on Demand” on page 90
- Section 4.8, “HMC Capacity on Demand menus” on page 98
- Section 4.9, “Capacity on Demand configuration rules” on page 102
- Section 4.10, “Software licensing considerations” on page 104
- Section 4.11, “Capacity on Demand feature codes” on page 107

4.1 Capacity on Demand overview

IT budgets no longer allow for an infrastructure that can accommodate occasional peaks in demand but whose resources otherwise lie idle. With IBM CoD, you can get the processing or storage resources you need, when you need them, on demand. Because costs vary with usage, CoD can provide a highly cost-efficient method of handling the usage peaks and valleys that occur in any business. This model can allow companies to quickly scale their IT infrastructures to meet dynamic application requirements without purchasing additional servers or storage devices.

Through unique CoD offerings, IBM @server products 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. The four types of CoD activation configurations are discussed within this chapter, 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 service 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, or later

4.2 What’s new in Capacity on Demand?

The p5-590 and p5-595 Capacity on Demand (CoD) systems can be configured with inactive processor and/or memory resources that can be enabled dynamically and non-disruptively to the system. The CoD offering provides flexibility and improved granularity in processor and memory upgrades and is a useful configuration option to support future growth requirements. CoD also eliminates the need to schedule downtime to install additional resources at a given point in time. Throughout this chapter the term *capacity* is used to refer to processor and memory resources intended for use in CoD configurations.

Dynamic processor and memory resource activation was introduced into the pSeries product line as the *Capacity Upgrade on Demand* (CUoD) offering and was first available on the pSeries 670 and pSeries 690 @server models. CUoD was offered in three types of activation formats. Since the initial pSeries release of CUoD, the p5-590 and p5-595 have enhanced previously existing CUoD functionality and expanded the CoD solutions portfolio. Table 4-1 on page 85 provides a side-by-side comparison of CoD features between the POWER4 and POWER5 high-end servers.

Table 4-1 CoD feature comparisons

pSeries 670 and pSeries 690	p5-590 and p5-595
► Introduced Capacity Upgrade on Demand (CUoD) for the addition of permanent processor & memory system resources ► Created On/Off Capacity on Demand (On/Off CoD) as a pre-pay debit activation plan for temporary processor resources ► Created Trial Capacity on Demand (Trial CoD) offering for one-time, no-cost trial activation of processor and memory resources for a maximum of 30 consecutive days	► Retained Capacity Upgrade on Demand (CUoD) for the addition of permanent processor & memory system resources ► Changed On/Off CoD to a post-pay, self-managed activation plan for temporary processor and/or memory resources; there are significant differences between the POWER4 and POWER5 definitions of On/Off CoD ► Introduced Reserve Capacity on Demand (Reserve CoD) configuration plan for “autonomic” temporary activation of pre-paid processor resources; this plan serves as an enhancement to the POWER4 On/Off CoD plan ► Trial CoD offering modified to provide Web-based enablement option for one-time, no-cost trial activation of processor and memory resources using activation codes ► Introduced Capacity BackUp

The information at the following URL briefly explains the CoD process and features for all IBM pSeries ⓔserver and IBM Total Storage products:

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

4.3 Preparing for Capacity on Demand

In this section, a discussion of how to include Capacity on Demand in your enterprise.

4.3.1 Step 1. Plan for future growth with inactive resources

Processor features on the p5-590 and p5-595 servers are, by default, all CoD inactive. When initially configuring a new server, the configuration process requires the selection of the amount of installed processor features and the amount of activations desired out of the total installed. For example, by configuring a 16-way processor node to be installed and choosing to activate 12 processors will allow for four processors to be available for future activation.

Software licensing charges do not apply to inactive processor resources. Hardware warranty is included on inactive resources and hardware maintenance charges are applicable at a reduced rate on CoD inactive resources.

Some memory features offered for the p5-590 and p5-595 servers offer 50 percent inactive CoD memory. For example, a feature may include 8 GB of total memory but only 4 GB of which are activated and available. The remaining 4 GB can be activated at a future time. Memory CoD is independent of processor CoD and can be configured with or without the use of processor CoD.

CoD inactive resources can also be ordered as upgrades to installed systems under the same methodology as when ordering the CoD features on a new system.

4.3.2 Step 2. Choose the amount and desired level of activation

Either permanent or temporary activation of resources can be ordered by purchasing the features associated with the type of activation desired. When an order for a permanent or pre-paid activation is received an activation code is generated specific to the ordered features and a letter containing the activation code or codes are mailed. In addition, the activation code is also available from the following Web site at the time the order processing is completed:

http://www-912.ibm.com/pod/pod

When received, the user can enter the activation code at the HMC console and the activated resources are then available.

> **Important:** Activation codes must be entered in sequential order for the same type of activation. Therefore, if multiple processor activations are ordered, the codes must be entered onto the server in chronological order depending on the date the code is generated.

There are no restrictions associated with the resale of a systems with inactive CoD resources. However, users of the On/Off CoD offering for Models 590 and 595 are required to terminate this function (at no charge) before the system is resold.

4.4 Types of Capacity on Demand

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

IBM has established four types of CoD offerings on the p5-590 and p5-595 systems, each with a specific activation plan. Providing different types of CoD offerings gives clients maximum 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 IBM 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 & memory
- On/Off Capacity on Demand (CoD) for processors & memory
- Reserve Capacity on Demand (CoD) for processors only
- Trial Capacity on Demand (CoD) for processors or memory

The p5-590 and p5-595 servers use a 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. The CoD activation process is addressed in Section 4.6, “Capacity on Demand activation procedure” on page 89, and the associated feature codes are summarized in Section 4.11, “Capacity on Demand feature codes” on page 107.

Table 4-2 provides a brief description of the four types of CoD offerings, identifies the proper name of the associated activation plan, 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 Global Financing representative. A functional overview of each CoD offering is provided in the subsequent sections.

Table 4-2 Types of Capacity on Demand (functional categories)

Activation plan	Functional category	Applicable system resources	Type of payment offering	Description
Capacity Upgrade on Demand (CUoD)	Permanent capacity for non-disruptive growth	Processor and memory resources	Pay when purchased	Provides a means of planned growth for clients who know they will need increased capacity but aren't sure when
On/Off Capacity on Demand (CoD)	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 (CoD)		Processor resources only	Pay before activation	
Trial Capacity on Demand (CoD)	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

4.5 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 offsite, 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 offers configurations of either a 32-way server (28 inactive and 4 active processors) or a 64-way server (60 inactive and 4 active processors).

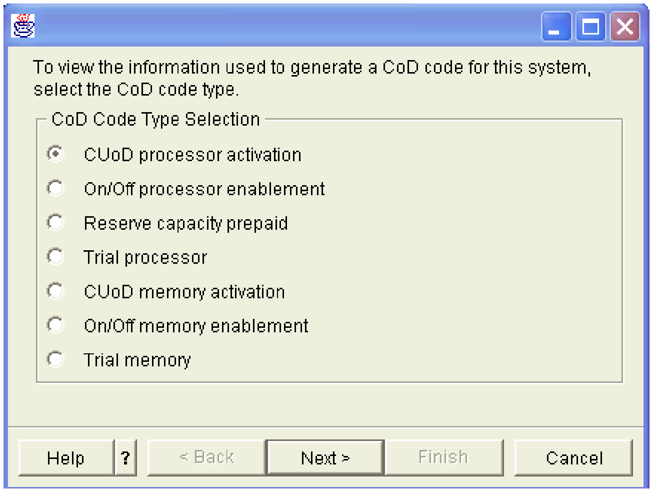
Capacity BackUp systems can be turned on at any time by using the On/Off CoD activation procedure for the needed performance during an unplanned system outage. Each Capacity BackUp configuration is limited to 450 On/Off CoD credit days per processor book. For clients who require additional capacity or 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 upon failover. When needed, HACMP can also activate DLPAR and CoD resources.

Capacity BackUp configurations:

- For p5-590: Capacity BackUp POWER5, 32 Standard Processors, 28 standby and 4 permanently active, along with 900 On/Off CoD credit processor days (450 On/Off CoD days per processor book). Select FC 7730 for this function.
- For p5-595: Capacity BackUp POWER5, 32 or 64 Standard Processors, 28 or 60 standby and 4 permanently active, along with 900 or 1800 On/Off CoD credit processor days (450 On/Off CoD days per processor book). Select FC 7731 for this function.
- For p5-595: Capacity BackUp POWER5, 32 or 64 Turbo Processors, 28 or 60 standby and 4 permanently active, along with 900 or 1800 On/Off CoD credit processor days (450 On/Off CoD days per processor book). Select FC 7732 for this function.

4.6 Capacity on Demand activation procedure

The p5-590 and p5-595 servers use an enablement code to activate CUoD, On/Off CoD, Reserve CoD, and Trial CoD resources. The activation process for IBM @server p5 CoD systems is quick and easy. The client simply places an order for a specific CoD feature code and supplies the necessary system configuration data when needed. The HMC order selection panel shown in Figure 4-1 is used to facilitate order placement.



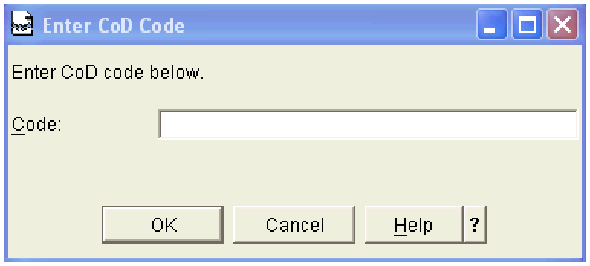
To view the information used to generate a CoD code for this system, select the CoD code type.

CoD Code Type Selection

● CUoD processor activation
○ On/Off processor enablement
○ Reserve capacity prepaid
○ Trial processor
○ CUoD memory activation
○ On/Off memory enablement
○ Trial memory

Figure 4-1 HMC Capacity on Demand Order Selection Panel

Once the order is placed, an encrypted key is sent to the client using the Web and by mail. This key is entered into the system using the hardware management console (HMC), which activates the desired number of additional processors or memory resources, as seen in Figure 4-2. There is no requirement to set up electronic monitoring of your configuration by IBM.


Enter CoD Code

Enter CoD code below.

Code:

Figure 4-2 Enter CoD Enablement Code (HMC Screen)

To order any of the CoD permanent or temporary activation offerings, see your sales channel representative. Feature codes for these activation offerings are summarized Section 4.11, “Capacity on Demand feature codes” on page 107.

The HMC will allow up to five invalid activation code entries. If more than five invalid activation codes are entered, the system must be rebooted before another activation code can be entered. Some types of failures are not counted against the five invalid attempts, including:

- User mis-typed activation code
- Previous activation code for the same system
- Activation code from a different system that contains a different processor type
- Activation code for memory from a different system is entered and this system does not contain any CUoD memory cards

4.7 Using Capacity on Demand

The following sections outline how to use each of the major CoD features.

4.7.1 Using Capacity Upgrade on Demand

Capacity Upgrade on Demand (CUoD), a derivative of CoD, was introduced on the pSeries 670 and pSeries 690 product models; this feature is still offered in the p5-590 and p5-595 systems. CUoD is intended to allow the addition of permanent processor or memory system resources for non-disruptive growth.

Additional processor and/or memory resources, also referred to as *capacity*, are designed into the original purchase as inactive resources, with minimal up-front charges or pricing premiums for this additional capacity.

The CUoD feature allows users to make permanent capacity upgrades rapidly without disrupting applications already running on the server. CUoD does not require any special contracts or monthly monitoring. However, once the resources are activated, they cannot be deactivated.

Capacity Upgrade on Demand activation procedure

IBM will provide the client with a 34-character encrypted key for permanent activation of CUoD resources, both processor and memory, on the p5-595. This key is entered into the system using the hardware management console (HMC) to activate the desired number of additional processors or memory resources. Upon entry of the activation code at the HMC, the system will have access to the additional processor(s) and memory. Processors must be activated in whole processor units; memory may be activated in 1 GB increments.

Both POWER5 processor cores are activated as part of the CUoD activation process. This results in optimized system performance and maintains consistent, predictable performance. A complex algorithm in the system firmware, based on factors such as memory groups, available inactive processors, and possible processor failure or deallocation situations, determines which pair of processors will be activated. Questions regarding either ordering or activation of inactive CUoD processors or memory, should be directed to your regular sales contact.

4.7.2 Using On/Off Capacity On Demand

On/Off Capacity on Demand (On/Off CoD) is a post-pay, self-managed offering that allows for the dynamic addition and removal of processor and memory capacity to p5 systems on a temporary basis. The ability to temporarily add capacity helps businesses cope with both predictable and unpredictable surges in transaction volume with the aid of temporary increases in capacity. These resources can be activated and deactivated quickly and efficiently as organizational demands dictate. In this temporary capacity model, the system owner only pays for the resources they have used, as capacity usage is monitored. This option provides a highly cost-effective strategy for handling seasonal or period-end fluctuations in activity and can enable you to deploy pilot applications without investing in new hardware.

Clients with pre-installed, inactive CUoD processor and memory resources can order temporary capacity using specific On/Off CoD feature codes. However, On/Off CoD has two distinct phases: enablement and activation. Once an On/Off CoD feature code is purchased, an enablement code will be supplied at the time of purchase to make the resources available for use. The administrator uses the enablement code to turn processors and memory on and off as desired, in increments of one processor or 1 GB of memory for 30 days of usage. These activations may be used for 30 consecutive days or turned on and off over a longer period of time.

The system monitors the amount and duration of the activations and generates a usage report which must be sent to IBM monthly. Billing for the activations is then based on the usage report on a quarterly basis. The processor and memory resources are paid for only after they are used, thereby allowing for a flexible, “pay as you go” plan.

Clients who wish to implement On/Off CoD features must sign an On/Off CoD contract. Additional information regarding activation of temporary On/Off resources is addressed in 4.6, “Capacity on Demand activation procedure” on page 89.

On/Off Capacity on Demand enablement procedure

On/Off CoD provides an innovative and flexible temporary processor and memory activation capability for planned or predictable peek processing requirements for p5-590 and p5-595 servers. It is important to note that On/Off CoD capacity must first be enabled before it can be activated and used. When an order is placed for an On/Off CoD enablement feature, the client receives an enablement key that allows the administrator to turn processors and memory on and off as desired, in increments of one processor and 1 GB of memory. The system monitors the amount and duration of the activations and generates a usage report which must be sent to IBM monthly. Billing for the activated resources is based on the usage report. Clients only pay for the processor days and memory days after they have been used.

The enablement process for On/Off CoD resources can be summarized in the following four steps:

1. Planning, registration, and contracts

   The client must ensure that the system meets the minimum levels for firmware, hardware management console (HMC), and operating system as outlined in Section , “When the pre-installed, inactive capacity becomes necessary, an administrator simply enters a code to activate the desired resources. The resource activation process is addressed in 4.6, “Capacity on Demand activation procedure” on page 89. Additionally, there are activation and configuration rules for processor and memory resource activation on the p5-590 and p5-595 systems. This information is summarized in 4.9, “Capacity on Demand configuration rules” on page 102.” on page 104. Once the system configuration and code levels have been verified, the client must access the On/Off COD Sales Channel Web site to register for On/Off CoD program participation and identify the target server(s) involved. Finally, the client must sign contracts and licensing agreements to establish the desired billing method. The contracts will be filed with the Temporary Capacity on Demand (TCoD) Project Office. The Sales Channel Contracts and Registration page can be accessed using the URL below:

`http://www-912.ibm.com/supporthome.nsf/document/28640809`

2. Enablement

   Before enabling temporary capacity onto the target server(s), a client must place an order for the appropriate enablement feature. The hardware management console (HMC) provides a wizard to generate an order for a CoD code (Figure 4-1 on page 89). The order information can be saved on a remote file system or to a DOS formatted floppy diskette. Once the order is placed, IBM Manufacturing will generate an enablement code and deliver it to the client. The client will receive a hard copy of the enablement code by mail. The enablement code will also be posted online to allow convenient retrieval access. The enablement code is used to activate the resources using the hardware management console (HMC) as seen in Figure 4-2 on page 90. An On/Off CoD enablement code enables the server to use up to 360 processor days that can be requested as temporary capacity. When the limit is reached, a new enablement feature must be ordered, and a new enablement code must be entered to activate the resource(s).

   Clients may order multiple On/Off CoD enablement features, if desired. However, separate features are needed for processor and memory enablement. The HMC controlling the p5-590 or p5-595 server will be used to set the parameters desired when using multiple On/Off CoD activation features.

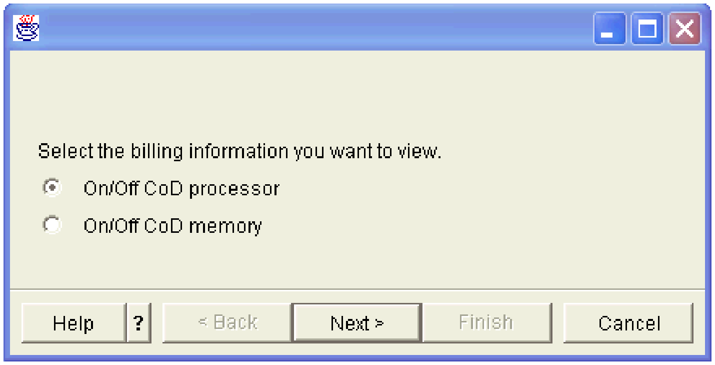
3. Usage

   The hardware management console (HMC) is used to turn capacity on and off as needed. The client is obligated to report usage once per month, either via the electronic Service Agent or manually.

4. Reporting and order processing and billing

   The client must report their on/off usage to IBM at least monthly. This monthly usage data will be stored in the TCoD database. This information is used to compute the billing data, which is provided to the sales channel on a quarterly basis. The sales channel places an order for the quantity of On/Off processor or memory used and invoices the client.

   The HMC has a billing wizard (Figure 4-3 on page 94) to provide the client with information related to On/Off CoD processor and memory billing information. This information can also be saved on a remote system or on a DOS formatted floppy diskette.



Select the billing information you want to view.

● On/Off CoD processor

○ On/Off CoD memory

Figure 4-3 HMC Billing Selection Wizard

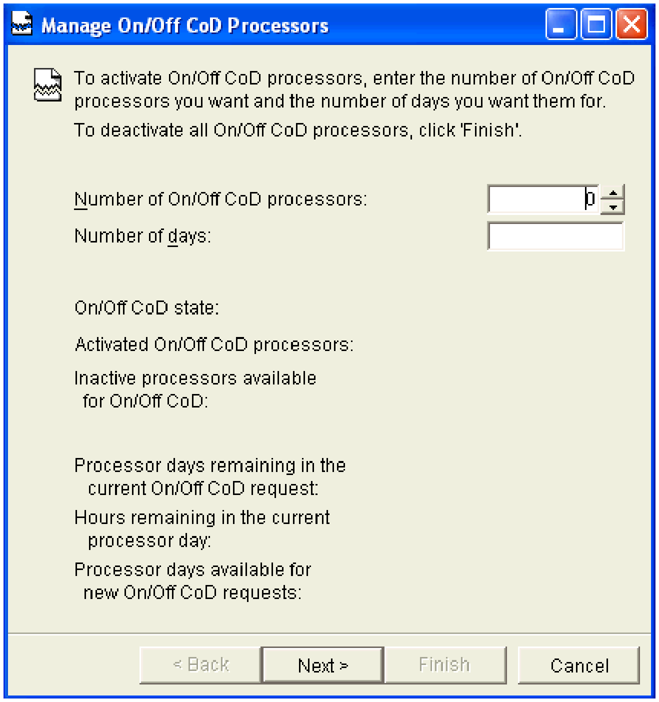
A request for temporary capacity will be counted in “resource days”, and the client will be required to report this request for temporary capacity in order to be billed for temporary capacity in the form of requested and/or unreturned resource days. Before the time expires on a given On/Off request, the resources must be returned to the unlicensed processor pool. If they are not returned, the resources will continue to be counted and billed as “unreturned resource days” as defined in the client’s contract. Additional On/Off requests will not be honored if the limit of resource days has been reached or unlicensed resources are in use on the machine.

The following examples of the usage calculation:

- A processor day is measured each time a processor is activated in a 24-hour period or any fraction thereof. If four processors are activated for two consecutive hours of testing or production, the result is four processor usage days.
- If the same client activates two of the previously activated processors in the same 24-hour period, the result is six processor usage days since a new measurement day begins each time processors are activated.

Managing On/Off Capacity on Demand resources

The hardware management console (HMC) is used to manage and enable resources temporarily on a p5-590 or p5-595 machine. After the On/Off enablement code has been obtained and entered, the client can request enablement of the desired number of On/Off CoD resources for the specified number of usage days. Each On/Off CoD activation ordered authorizes activation of one processor or 1 GB of memory for 30 days of usage. Each individual On/Off CoD activation is valid for a maximum of 192 days. Figure 4-4 provides a snapshot of the HMC activation screen used to define the number of processor resources and days. The activation screen for memory resources is almost identical.



Manage On/Off CoD Processors

To activate On/Off CoD processors, enter the number of On/Off CoD processors you want and the number of days you want them for.

To deactivate all On/Off CoD processors, click 'Finish'.

Number of On/Off CoD processors: 0

Number of days:

On/Off CoD state:

Activated On/Off CoD processors:

Inactive processors available for On/Off CoD:

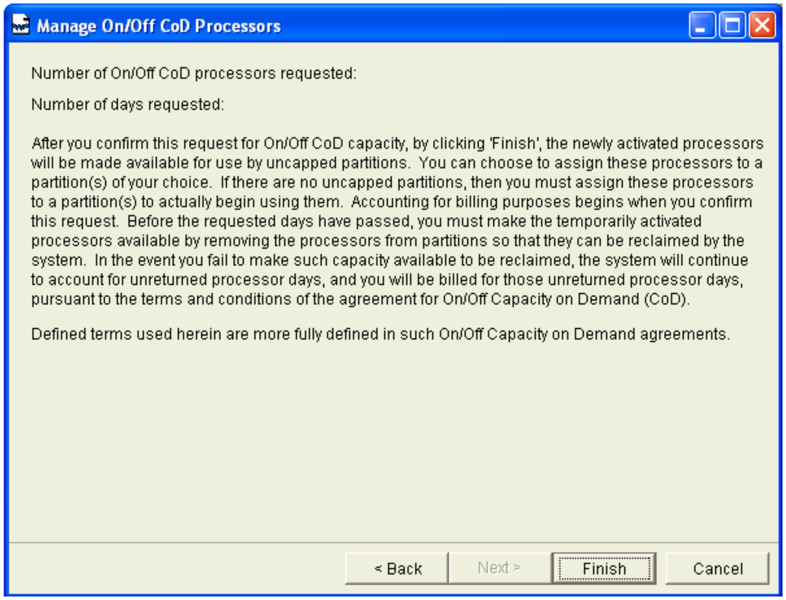
Processor days remaining in the current On/Off CoD request:

Hours remaining in the current processor day:

Processor days available for new On/Off CoD requests:

Figure 4-4 Manage On/Off CoD Processors HMC Activation Screen

Once the user has selected the desired amount and usage days for the On/Off capacity, a summary will be presented as seen in Figure 4-5 on page 96. This summary screen and associated legal text will only be presented when activating On/Off CoD resources; it will not be displayed when the resources are deactivated. The confirmation panel for managing On/Off CoD memory resources is almost identical.



Number of On/Off CoD processors requested:

Number of days requested:

After you confirm this request for On/Off CoD capacity, by clicking 'Finish', the newly activated processors will be made available for use by uncapped partitions. You can choose to assign these processors to a partition(s) of your choice. If there are no uncapped partitions, then you must assign these processors to a partition(s) to actually begin using them. Accounting for billing purposes begins when you confirm this request. Before the requested days have passed, you must make the temporarily activated processors available by removing the processors from partitions so that they can be reclaimed by the system. In the event you fail to make such capacity available to be reclaimed, the system will continue to account for unreturned processor days, and you will be billed for those unreturned processor days, pursuant to the terms and conditions of the agreement for On/Off Capacity on Demand (CoD).

Defined terms used herein are more fully defined in such On/Off Capacity on Demand agreements.

Figure 4-5 Manage On/Off CoD HMC Confirmation Panel and Legal Statement

4.7.3 Using Reserve Capacity on Demand

Reserve Capacity on Demand (Reserve CoD) helps businesses meet unexpected and intermittent peeks in processing requirements in a shared dynamic LPAR environment. This plan is designed to allow p5-590 and p5-595 servers to have optimized, automatically managed temporary activation of CoD processor resources.

The user purchases a block of 30 Processor Days of usage time and then assigns inactive processors to the shared processor pool. When these inactive processors are activated, all partitions in the shared processor pool will start executing on all the processors in the shared processor pool, increasing application throughput. The server automatically manages the workload and only charges against the pre-paid Processor Day account when the workload exceeds 100 percent of the base permanently activated processors in the shared processing pool by 1/10 of a processor unit.

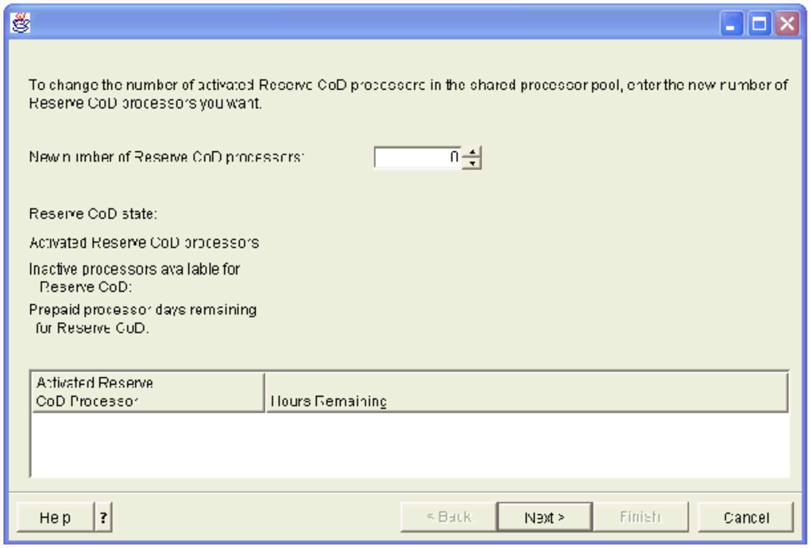
Processor resources enabled using the Reserve CoD feature are active for up to 30 days as required at the client's discretion without further contracts or authorization. Prepaid reserve capacity is supported by the Advanced POWER Virtualization feature. Reserve CoD is also referred to as metered Capacity on Demand.

Reserve Capacity on Demand activation procedure

Reserve CoD allows clients to implement on/off processor resources using a prepaid, automated plan. The 30 Days Prepaid Reserve Capacity feature allows use of processors for up to 30 days as required at the client's discretion without further contracts or authorization. The additional capacity is enabled using the appropriate Reserve CoD feature code.

Reserve CoD allows p5-590 and p5-595 servers to have optimized, automatically managed temporary activation of CoD processors. The user purchases a block of 30 Processor Days of usage time and then assigns inactive processors to the shared processor pool. When these inactive processors are activated, all partitions in the shared processor pool will start executing on all the processors in the shared processor pool, increasing application throughput. The server then automatically manages the workload and only charges against the processor day account when the workload exceeds 100% of the base, permanently activated, processors in the shared processing pool by 1/10th of a processor unit.

The HMC has a panel for Reserve CoD processor information that also allows the user to activate or deactivate Reserve CoD resources.



To change the number of activated Reserve CoD processors in the shared processor pool, enter the new number of Reserve CoD processors you want.

New number of Reserve CoD processors: 0

Reserve CoD state:

Activated Reserve CoD processors

Inactive processors available for Reserve CoD:

Prepaid processor days remaining for Reserve CoD.

Activated Reserve CoD Processor	Hours Remaining

Figure 4-6 HMC Reserve CoD Processor Activation Screen

4.7.4 Using Trial Capacity on Demand

Trial Capacity on Demand (Trial CoD) allows businesses to evaluate the use of inactive processors and memory for up to 30 continuous days at no charge. Trial

CoD is enabled by registering at the pSeries CoD Web site and electronically obtaining a Trial CoD activation code, as described in Section , “Trial Capacity on Demand activation procedure” on page 98.

System owners can activate additional processing or memory capacity, up to the maximum number of processors and memory installed on the server. Ideal conditions for enabling Trial CoD include testing new applications, handling workload peaks and testing, determining the effect of added computing resources on existing workloads, or supporting your business while in the process of purchasing permanent capacity upgrades.

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.

Trial Capacity on Demand activation procedure

Owners of @server p5 and pSeries systems with CUoD capabilities can activate all available inactive CUoD processors and memory resources one time, for up to 30 contiguous days, at no charge. Trial CoD for pSeries p5-590 and p5-595 servers is enabled by registering at the pSeries CoD Web site and electronically receiving an activation key (Figure 4-2 on page 90).

The pSeries CoD registration Web site can be accessed using the following URL

https://www-912.ibm.com/tcod_reg.nsf/TrialCod?OpenForm

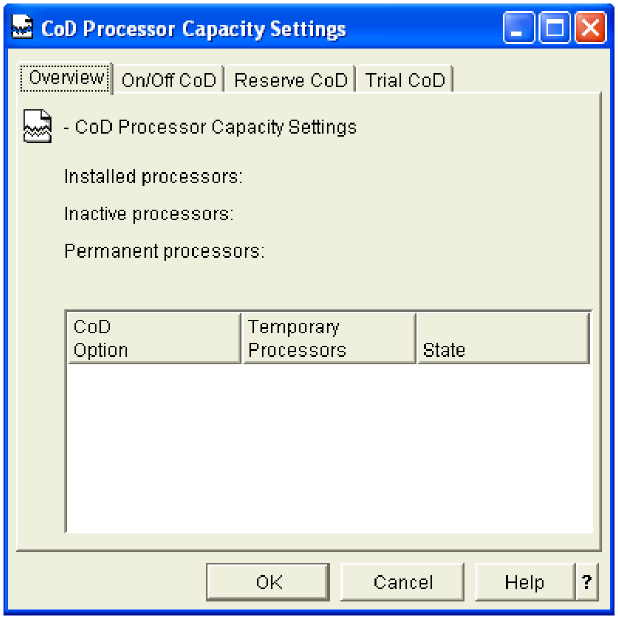
The client is allowed one use of Trial CoD when the system initially boots or after the client has purchased CUoD resources which were previously activated through the Trial CoD function. The additional temporary resources can be configured as permanent resources at a later date using the CUoD activation code or returned back to the system as CUoD resources. An additional 30-day trial period is available each time a client purchases a permanent processor activation, but the activated resources on these subsequent trial periods are limited to up to two processors and 4 GB of memory, if available on the server. This offering allows clients to activate additional resources temporarily for benchmarking, testing, or immediate use while waiting for the purchase of a CoD activation feature.

4.8 HMC Capacity on Demand menus

The hardware management console (HMC) provides a series of CoD menus to allow the user to observe and manage their CoD configurations. These CoD menus are accessible by using the ‘**Selected**’ menu from the HMC menu bar. The ‘**Selected**’ menu will have the following structure for CoD enabled systems:

- `Property`
- `Delete`
- `Create >`
- `Capacity On Demand >`
  - `Enter CoD Code`
  - `Processor >`
    - `Capacity Settings`
    - `Manage On/Off CoD`
    - `Manage Reserve CoD`
    - `Stop Trial CoD`
    - `Shared Pool Utilization`
  - `Memory >`
    - `Capacity Settings`
    - `Manage On/Off CoD`
    - `Stop Trial CoD`
  - `Show History Log`
  - `Show Code Information`
  - `Show Billing Information`

The HMC also has a set of panels that summarize the existing CoD processor and memory capacity settings. The following figures (Figure 4-7, Figure 4-8 on page 100, Figure 4-9 on page 101, and Figure 4-10 on page 101) provide examples of the CoD capacity settings for processors; the screens for memory capacity settings are almost identical.



- CoD Processor Capacity Settings

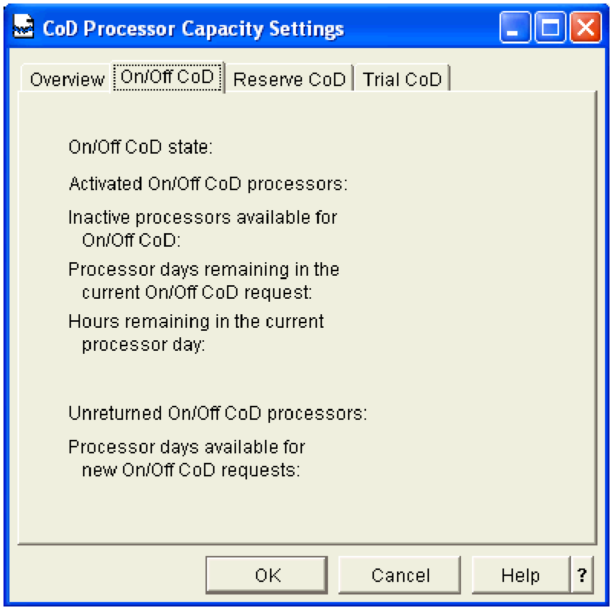
Installed processors:

Inactive processors:

Permanent processors:

CoD Option	Temporary Processors	State

Figure 4-7 CoD Processor Capacity Settings Overview HMC screen



On/Off CoD state:

Activated On/Off CoD processors:

Inactive processors available for On/Off CoD:

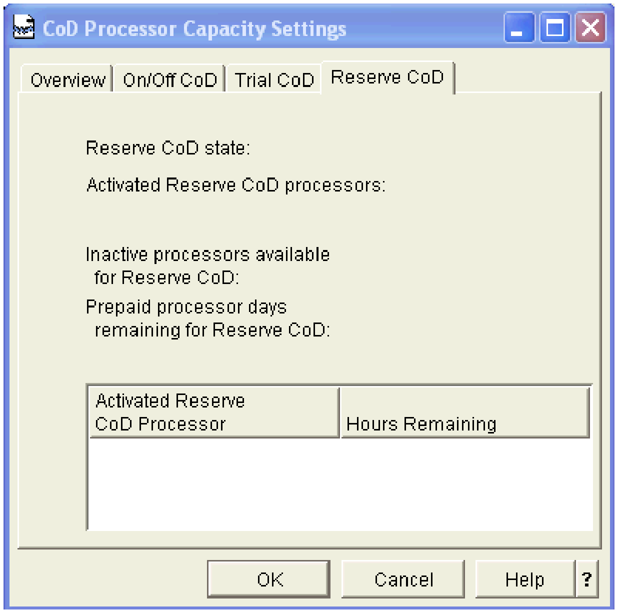
Processor days remaining in the current On/Off CoD request:

Hours remaining in the current processor day:

Unreturned On/Off CoD processors:

Processor days available for new On/Off CoD requests:

Figure 4-8 CoD Processor Capacity Settings On/Off CoD HMC screen



Reserve CoD state:

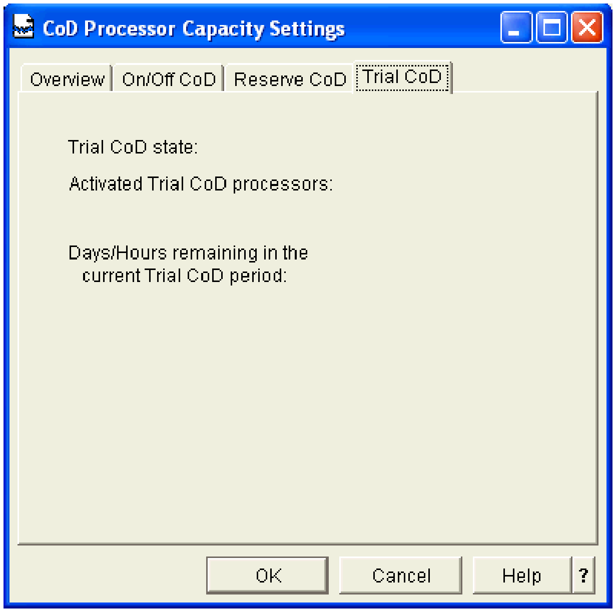
Activated Reserve CoD processors:

Inactive processors available for Reserve CoD:

Prepaid processor days remaining for Reserve CoD:

Activated Reserve CoD Processor	Hours Remaining

Figure 4-9 CoD Processor Capacity Settings Reserve CoD HMC screen



Trial CoD state:

Activated Trial CoD processors:

Days/Hours remaining in the current Trial CoD period:

Figure 4-10 CoD Processor Capacity Settings “Trial CoD” HMC Screen

4.8.1 HMC command line functions

In addition to the GUI interface to Capacity on Demand functions on the Hardware Management Console, command line functions are available for most actions. See the HMC documentation regarding use of remote command line functions.

4.9 Capacity on Demand configuration rules

All p5-590 and p5-595 multichip module (MCM) processor books are initially implemented as 16-way CUoD offerings with zero active processors. All p5 platforms are shipped with a CUoD anchor card. This card stores information regarding system CoD, including the activation entitlements (once entered) and data regarding capacity usage.

4.9.1 Processor capacity Upgrade on Demand configuration rules

As previously mentioned, all processor books available on the p5-590 and p595 are initially implemented as 16-way CUoD offerings with zero active processors. A minimum of eight permanently activated processors are required on the p5-590 server. A minimum of sixteen permanently activated processors are required on the p5-595 server. The number of permanently activated processors is based on the number of processor books installed, and is summarized in Table 4-3.

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

Table 4-3 Permanently activated processors by MCM

p5-590	p5-595
► One processor book (MCM) installed requires system to have 50% active processors. ► Two processor books (MCM) installed requires system to have 50% active processors.	► One processor book (MCM) installed requires system to have 50% active processors. ► Two processor books installed requires system to have 50% active processors. ► Three processor books installed requires system to have 50% active processors. ► Four processor books installed requires system to have 50% active processors.

4.9.2 Memory capacity Upgrade on Demand configuration rules

CUoD for memory may be used in any available memory position. Additional CUoD 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.

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

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

4.9.3 Trial Capacity on Demand configuration rules

Trial CoD for pSeries p5-590 and p5-595 servers is enabled by registering at the pSeries CoD Web site and electronically receiving an activation key. When clients purchase permanent processor activation on p5-590 and p5-595 servers, they have another 30 contiguous day trial. These subsequent trials are limited to the activation of 2 processors and 4 GB of memory.

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.

4.9.4 Dynamic processor sparing

The p5-590 and p5-595 keeps a pool of spare, unused processors for dynamic processor allocation. CUoD processors that have not been activated are allocated to this dynamic processor sparing pool. If the server detects an impending failure of an active processor, it will attempt to activate one of the unused CUoD processors from the pool 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. The processor pool is managed by the POWER Hypervisor and the operating system.

When the pre-installed, inactive capacity becomes necessary, an administrator simply enters a code to activate the desired resources. The resource activation process is addressed in 4.6, “Capacity on Demand activation procedure” on page 89. Additionally, there are activation and configuration rules for processor and memory resource activation on the p5-590 and p5-595 systems. This information is summarized in 4.9, “Capacity on Demand configuration rules” on page 102.

4.10 Software licensing considerations

Logical partitioning (LPAR) and Capacity on Demand (CoD) are important elements in the Virtualization capabilities of IBM @server, which in turn are a key component of the IBM On Demand strategy. Virtualization provides clients with the flexibility to easily and dynamically adapt their infrastructure to fluctuations in their workloads, and Capacity on Demand provides the flexibility to allow clients non-disruptive growth and be able to enable additional processors and memory temporarily for unplanned and unexpected peek load requirements. Software licensing, for the operating system and for IBM middle ware as well as for ISV applications is also an important consideration, and IBM has provided sub-capacity licensing and daily license entitlements to give clients the flexibility to choose how they will license the operating systems and selected middle ware products to help lower the Total Cost of Operations. This section provides current licensing rules for IBM operating systems and IBM middleware.

4.10.1 License entitlements for permanent processor activations

Clients license an IBM operating system (for example, IBM i5/OS or IBM AIX) for a machine, and they purchase license entitlements based on the number of processors in a dynamic LPAR or the shared processor pool. With AIX 5L on POWER5 servers, sub-capacity licensing is allowed. That requires clients to purchase license entitlements only for the total number of processors that AIX 5L or selected IBM middle ware is executing on. For example, if a client has a 16-way server and is running AIX 5L applications on 12 processors and one of the LINUX distributions on the remaining four processors, the client must have 12 license entitlements of AIX. There are just a few rules to determine the number of license entitlements required, depending on the environment.

Dedicated partition environment

In the dedicated partition environment, the rules require that the Dynamic LPAR be defined in increments of whole processors so the number of license entitlements is equal to the number of processors dedicated to the Dynamic LPAR.

Shared partition environment

In the Shared Pool environment, Dynamic LPARs may be defined with less than a full processor entitlement. IBM does not license software products at the sub-processor level even though Advanced POWER Virtualization allows clients to define a dynamic LPAR with less than a full processor entitlement. For software licensing, the client must round up to the next whole number to determine how many license entitlements are required for the OS and middle ware executing in the partition. For example a shared partition may be defined with a 0.5 processor entitlement. However, for software programs running in this 0.5 processor partition, either operating system or middleware, the client must purchase a full license entitlement. There is also a difference in calculating the entitlements for capped and uncapped partitions in the shared environment. With capped partitions, the partition is defined with a max processor unit value (PRU) which defines the maximum number of processors units (virtual processors) the partition may consume. This value is the one used to determine how many license entitlements are required for the partition.

As an example, if a capped partition is defined with PRU = 1.5 then the number of license entitlements is rounded up to the next whole number and is equal to two license entitlements. With Uncapped partitions, the partition is defined with a max virtual processor (VP) value which defines the maximum number of physical processors within the share processor pool which may be concurrently executing on behalf of the partition.

This value is the one used to determine how many license entitlements are required for the partition. As an example, if an uncapped partition is defined with VP = 7 where the shared processor pool has 12 physical processors assigned, then the number of software license entitlements required for this uncapped partition is seven entitlements.

Determining the number of license entitlements required

The client must add the sum of the number of processors in each partition running the OS or middleware after rounding up in each partition where the OS or the middleware product is running. This total equals the total number of software license entitlements that the client must purchase up to the total number of active processors in the server.

As an example, if a client has a 32-way server with 24 active processors, 8 inactive processors, and 20 active processors assigned to the Shared Processor Pool, and had defined partitions in the following manner (Table 4-4):

Table 4-4 License entitlement example

Partition	Type	Processors assigned	SW entitlements
A	Dedicated	Assigned 1 Processor	1 Entitlement
B	Dedicated	Assigned 2 Processors	2 Entitlements
C	Dedicated	Assigned 1 Processor	1 Entitlements

Partition	Type	Processors assigned	SW entitlements
SA	Shared/Capped	PRU = 1.5 Processors	2 entitlements
SB	Shared/Capped	PRU = 2.5 Processors	3 entitlements
SC	Shared/Uncapped	VP = 12 Processors	12 entitlements
SD	Shared/Uncapped	VP = 12 Processors	12 entitlements

When adding up the total number of required software entitlements, there are a total of four required for the dedicated partitions, and a total of 29 required for the shared partitions, giving a total number of entitlements required equal to 33.

However, the server has a total of 24 active processors so the total number of software license entitlements is 24. Clients do not have to have more license entitlements than the total number of active processors in the server.

4.10.2 License entitlements for temporary processor activations

When clients use On/Off CoD, Reserve CoD, or Trial CoD, they must purchase software license entitlements for each processor day utilized to be compliant with their International Program License Agreements for processor-based software programs. Clients have the choice of purchasing an additional full license of the OS or middle ware, which will give them 365 processor days of license entitlements per year, or purchase individual daily software license entitlements for just the number of processor days utilized.

Daily license entitlements may only be purchased and used under the following conditions:

- The client must have purchased at least one full license of the OS or middle ware.
- The client must have a current On/Off CoD, Reserve CoD, or Trial CoD feature in effect on the server (daily license entitlements may only be used in conjunction with temporary processor activations on the server).
- There is no software defect or upgrade support associated with a daily license entitlement, if the client wants defect or upgrade support for the OS or middle ware, they must have a current SWMA on the base license of the OS or middle ware.
- Clients must sign contract Z125-6907, Amendment for iSeries and pSeries Software On/Off Capacity on Demand.

An example is if the client has a 32-way server with 24 active processors and 8 inactive processors with AIX 5L, Cluster Systems Management (CSM), and HACMP running in the server, and purchases an On/Off CoD enablement feature. When the client enters the enablement key into the HMC, processors and memory may be activated and deactivated when needed. Quarterly, the client will receive a report from IBM indicating how many processor days have been utilized. This data is captured by the system and is reported monthly to IBM. If the report shows 30 processor days were utilized in the quarter, the client must then order 30 daily license entitlements for AIX 5L, and if CSM and HACMP were also running on these temporary processors, the client must order 30 daily license entitlements for CSM and HACMP.

4.11 Capacity on Demand feature codes

The CoD feature (enablement) codes used to order CoD capabilities on the p5-590 and p5-595 are summarized in Table 4-5.

Table 4-5 p5-590 and p5-595 CoD Feature Codes

Inactive resource feature		**CoD feature codes**			
Description	**Feature code (FC)**	**Permanent activation CUoD**	**Reserve CoD**	**Processor**	**Memory**
				On/Off Enablement / Billing	
p5-590					
0/16 Processors (1.65 GHz POWER5)	FC 7981	FC 7925	FC 7926	FC 7839 / FC 7993	
2/4 GB DDR1 Memory	FC 7816	FC 7970			FC 7973 / FC 7974
4/8 GB DDR1 Memory	FC 7835	FC 7970			FC 7973 / FC 7974
p5-595					
0/16 Processors (1.65 GHz POWER5)	FC 7988	FC 7990	FC 7991	FC 7994 / FC 7996	
0/16 Processors (1.9 GHz POWER5)	FC 7813	FC 7815	FC 7975	FC 7971 / FC 7972	

Inactive resource feature		CoD feature codes			
Description	Feature code (FC)	Permanent activation CUoD	Reserve CoD	Processor	Memory
				On/Off Enablement / Billing	
2/4 GB DDR1 Memory	FC 7816	FC 7970			FC 7973 / FC 7974
4/8 GB DDR1 Memory	FC 7835	FC 7970 FC 7799*			FC 7973 / FC 7974
* FC 7799 enables 256 1 GB Memory Activations (at one time) for FC 7835 on p5-595 only.					

5

Configuration tools and rules

This chapter is intended to help IBM employees and IBM Business Partners prepare a valid p5-590 and p5-595 configuration to meet a client’s requirement.

- Section 5.1, “Configuration tools” on page 110 introduces the IBM Configurator for e-business, also known as e-config and the LPAR Validation Tool
- Section 5.2, “Configuration rules for p5-590 and p5-595” on page 115 details server configuration rules for the p5-590 and p5-595
- Section 5.3, “Capacity planning considerations” on page 133 outlines considerations regarding capacity and performance

5.1 Configuration tools

IBM provides the IBM Configurator for e-business, also known as *e-config*, to help IBM employees and IBM Business Partners properly and efficiently configure their IBM @server platforms. The purpose of this section is to summarize capabilities and features of the IBM Configurator tool.

5.1.1 IBM Configurator for e-business

The IBM Configurator tool, e-config, is an application that provides configuration support for hardware, software, and peripherals associated with IBM product lines that are available for purchase. The IBM Configurator tool allows you to perform the following functions:

- Configure and upgrade IBM @server systems and subsystems
- Configure multiple product lines with just one tool
- Presents you only with the screens you need to build your configuration, rather than all product categories and configuration options
- Allows you to view all of your selections, from a high level, without moving backward
- View list prices as the configuration is being constructed
- Use system diagrams to review expansion options and explore configuration alternatives
- Provides reassurance that the configuration components will integrate together for an optimal solution

Models supported

The IBM Configurator for e-business can design solutions using the following components:

- IBM @server iSeries and AS/400®
- IBM @server p5
- IBM @server pSeries and RS/6000®
- IBM @server zSeries and S/390®
- Networking Products
- Printers
- Retail Store Solutions
- Storage Products

Note: RS/6000 SP configurations are supported by PCRS6000.

For information on how to get e-config, check the e-config home page:

http://ftp.ibmlink.ibm.com/econfig/announce/index.htm

e-config for p5-590 and p5-595

The IBM Configurator allows you to make valid configurations for p5-590 and p5-595 servers. The tool also checks for prerequisite components and features, as well as identifies incompatible components. It will often times correct such configuration errors automatically. In order to efficiently construct a customized platform configuration, the user should have an understanding of the configuration limitations, such as the minimum system resources, inherent to the p5-590 and p5-595 server architecture prior using the e-config tool. These considerations are outlined in the sections that follow.

5.1.2 LPAR Validation Tool

The LPAR Validation Tool (LVT) is a tool that will assist you in validating the resources assigned to LPARs. The LVT was designed specifically for the latest p5 servers. As partitioned environments grow increasingly complex, the LVT tool should be your first resource to determine how a system can be effectively partitioned.

LVT is a Java-based tool that is loaded on a Microsoft Windows 95 or above workstation with at least 128 MB of free memory. Its footprint on disk, at the time of writing, is about 47 MB. It includes an IBM Java Runtime Environment 1.4. The installation adds an icon to the desktop.

For information, including user’s guide and download information, see:

http://www.ibm.com/servers/eserver/iseries/lpar/systemdesign.htm

Questions and comments can be sent to:

mailto:rchlvt@us.ibm.com

During the development of this publication, the LVT was receiving regular updates. Installation only required a couple of minutes, with the most time devoted to downloading the code. An update, available as a separate file, brings the base code to the latest level and is significantly smaller in size.

System Selection dialog

Once the tools is launched you can create a new configuration or load an existing one, as shown in Figure 5-1 on page 112.

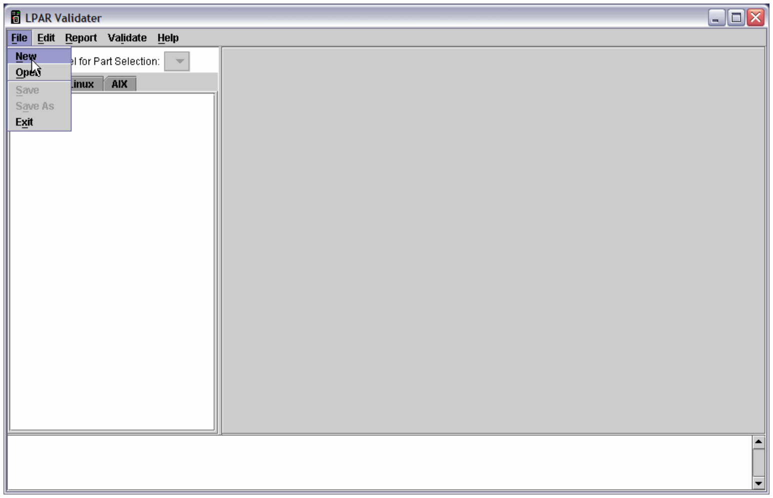


Figure 5-1 LPAR Validation Tool - creating a new partition

Upon opening up a new configuration, the resulting dialog (Figure 5-2 on page 112) provides places to select the basic attributes of the machine you plan to validate.

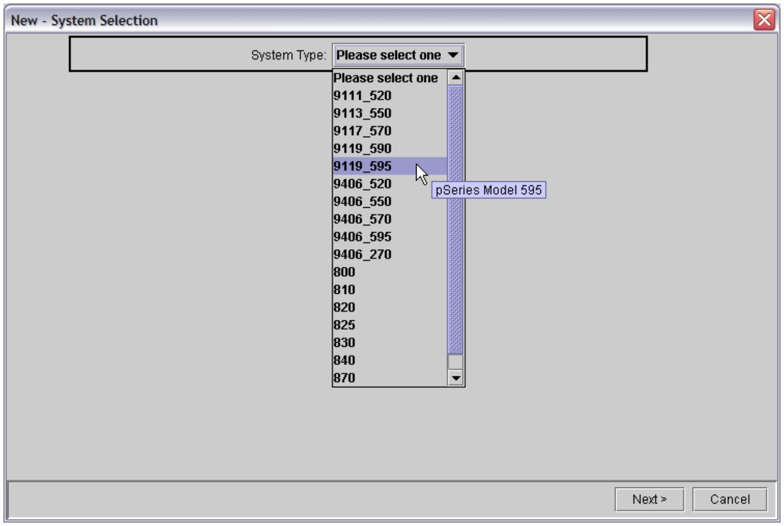
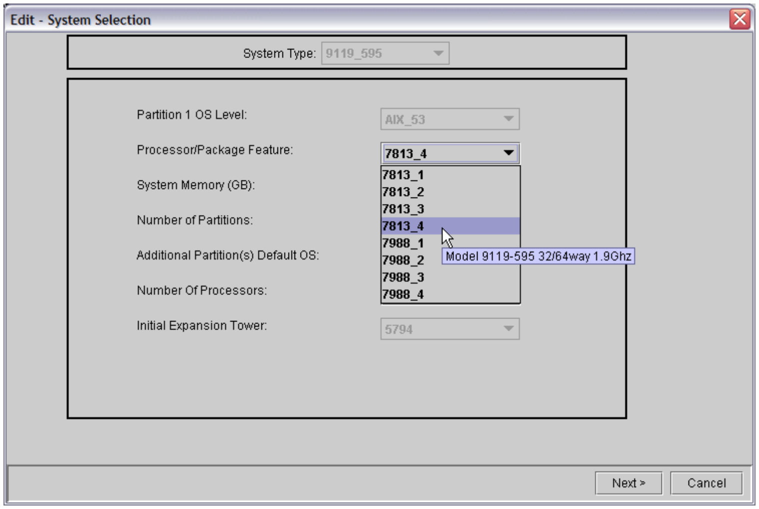


Figure 5-2 LPAR Validation Tool - System Selection dialog

Hold your cursor over a field, and additional information is provided, as shown in (Figure 5-3 on page 113).



System Type: 9119_595

Partition 1 OS Level: AIX_53

Processor/Package Feature: 7813_4

System Memory (GB):

Number of Partitions:

Additional Partition(s) Default OS:

Number Of Processors:

Initial Expansion Tower: 5794

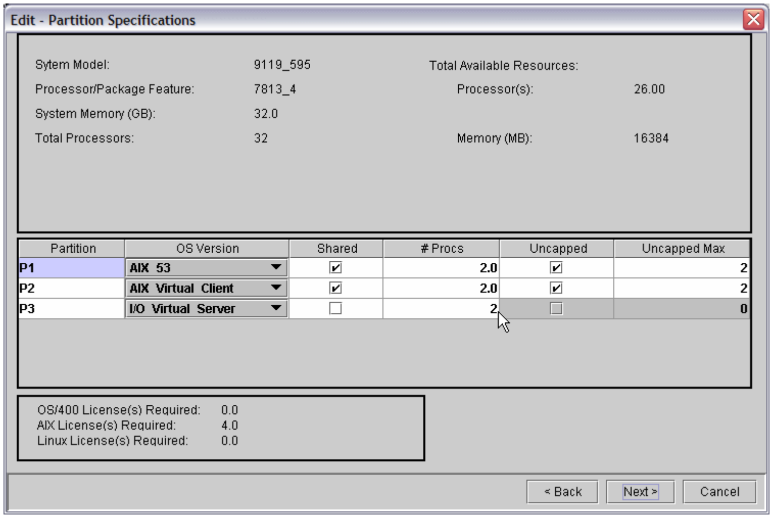
7813_1
7813_2
7813_3
7813_4
7988_1
7988_2
7988_3
7988_4

Model 9119-595 32/64way 1.9Ghz

Figure 5-3 LPAR Validation Tool - System Selection processor feature selection

Partition Specification dialog

In the next dialog, partition specifications are entered (Figure 5-4).



Sytem Model:	9119_595	Total Available Resources:		
Processor/Package Feature:	7813_4		Processor(s):	26.00
System Memory (GB):	32.0			
Total Processors:	32		Memory (MB):	16384

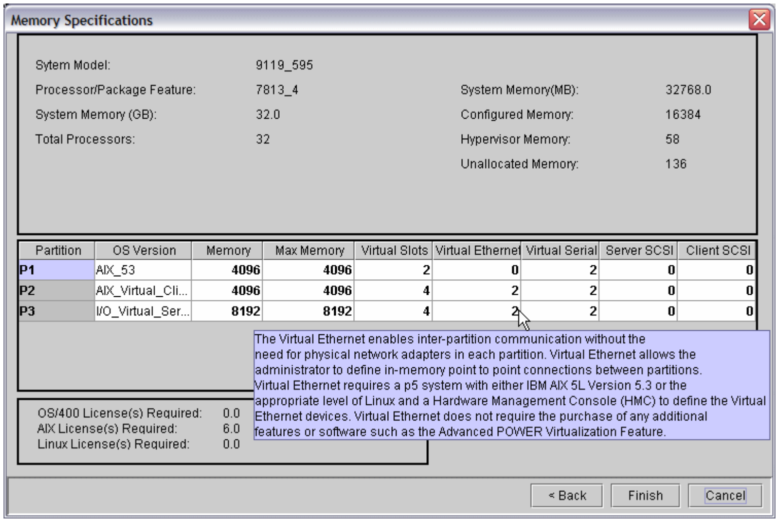
Partition	OS Version	Shared	# Procs	Uncapped	Uncapped Max
P1	AIX_53	☑	2.0	☑	2
P2	AIX_Virtual_Client	☑	2.0	☑	2
P3	I/O_Virtual_Server	☐	2	☐	0

OS/400 License(s) Required:	0.0
AIX License(s) Required:	4.0
Linux License(s) Required:	0.0

Figure 5-4 LPAR Validation Tool - Partition Specifications dialog

Memory Specification dialog

After the Partition Specification fields are complete, the memory specifications are entered for each of the logical partitions you previously specified (Figure 5-5).



Sytem Model:	9119_595		
Processor/Package Feature:	7813_4	System Memory(MB):	32768.0
System Memory (GB):	32.0	Configured Memory:	16384
Total Processors:	32	Hypervisor Memory:	58
		Unallocated Memory:	136

Partition	OS Version	Memory	Max Memory	Virtual Slots	Virtual Ethernet	Virtual Serial	Server SCSI	Client SCSI
P1	AIX_53	4096	4096	2	0	2	0	0
P2	AIX_Virtual_Cli...	4096	4096	4	2	2	0	0
P3	I/O_Virtual_Ser...	8192	8192	4	2	2	0	0

The Virtual Ethernet enables inter-partition communication without the need for physical network adapters in each partition. Virtual Ethernet allows the administrator to define in-memory point to point connections between partitions. Virtual Ethernet requires a p5 system with either IBM AIX 5L Version 5.3 or the appropriate level of Linux and a Hardware Management Console (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.

OS/400 License(s) Required:	0.0
AIX License(s) Required:	6.0
Linux License(s) Required:	0.0

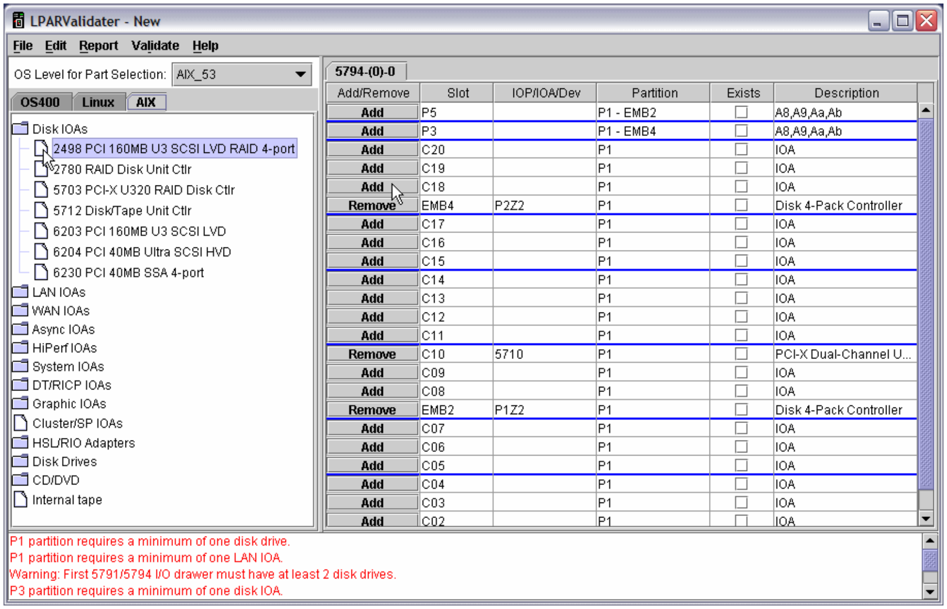
Figure 5-5 LPAR Validation Tool - Memory Specifications dialog

As the memory specifications are entered, the unallocated memory and the amount required by the hypervisor are shown. These values increase as the number of virtual devices defined increase. Internal validation prevents configurations that do not have enough resource.

This tool answers a common question in planning, “What is the memory usage of the POWER Hypervisor.”

LPAR Validation dialog

The final dialog enables you to assign features to the various slots defined, as shown in Figure 5-6.



OS Level for Part Selection: AIX_53

- Disk IOAs
  - 2498 PCI 160MB U3 SCSI LVD RAID 4-port
  - 2780 RAID Disk Unit Ctlr
  - 5703 PCI-X U320 RAID Disk Ctlr
  - 5712 Disk/Tape Unit Ctlr
  - 6203 PCI 160MB U3 SCSI LVD
  - 6204 PCI 40MB Ultra SCSI HVD
  - 6230 PCI 40MB SSA 4-port
- LAN IOAs
- WAN IOAs
- Async IOAs
- HiPerf IOAs
- System IOAs
- DT/RICP IOAs
- Graphic IOAs
- Cluster/SP IOAs
- HSL/RIO Adapters
- Disk Drives
- CD/DVD
- Internal tape

Add/Remove	Slot	IOP/IOA/Dev	Partition	Exists	Description
Add	P5		P1 - EMB2	☐	A8,A9,Aa,Ab
Add	P3		P1 - EMB4	☐	A8,A9,Aa,Ab
Add	C20		P1	☐	IOA
Add	C19		P1	☐	IOA
Add	C18		P1	☐	IOA
Remove	EMB4	P2Z2	P1	☐	Disk 4-Pack Controller
Add	C17		P1	☐	IOA
Add	C16		P1	☐	IOA
Add	C15		P1	☐	IOA
Add	C14		P1	☐	IOA
Add	C13		P1	☐	IOA
Add	C12		P1	☐	IOA
Add	C11		P1	☐	IOA
Remove	C10	5710	P1	☐	PCI-X Dual-Channel U...
Add	C09		P1	☐	IOA
Add	C08		P1	☐	IOA
Remove	EMB2	P1Z2	P1	☐	Disk 4-Pack Controller
Add	C07		P1	☐	IOA
Add	C06		P1	☐	IOA
Add	C05		P1	☐	IOA
Add	C04		P1	☐	IOA
Add	C03		P1	☐	IOA
Add	C02		P1	☐	IOA

```
P1 partition requires a minimum of one disk drive.
P1 partition requires a minimum of one LAN IOA.
Warning: First 5791/5794 I/O drawer must have at least 2 disk drives.
P3 partition requires a minimum of one disk IOA.
```

Figure 5-6 LPAR Validation Tool - slot assignments

From this screen a detailed report is available, and a validation engine can be selected to point out any errors in the configuration. If changes to the memory configuration are required, you can edit the configuration and change the values in error. When the configuration has been validated without error, you can feel confident that the resources selected will provide the configuration desired. At this point, if you choose, you can configure the system in an ordering tool, or MES upgrade an existing machine with the additional resources required to handle a new LPAR configuration.

5.2 Configuration rules for p5-590 and p5-595

In order to properly configure a p5-590 or p5-595, a minimum set of requirements and constraints must be satisfied. The purpose of this section is to explain those requirements and identify components that are needed to establish the base (minimum) platform configuration.

Besides the accuracy of the configuration, it is also important to configure the system that best meets the needs of the application. Individual system application requirements such as availability, performance, or flexibility, may require additional components or devices to produce an optimal system configuration.

> **Note:** Changes or enhancements to p5-590 and p5-595 may occur so that information in this chapter might not be completely accurate over time. Refer to the latest p5-590 and p5-595 sales manual for detailed information.

5.2.1 Minimum configuration for the p5-590 and p5-595

The purpose of this section is to establish the minimum configuration for a p5-590 and p5-595. Appropriate feature codes for each system component are also provided. The IBM Configurator tool will also identify the feature code for each component used to build your system configuration.

> **Note:** Throughout this chapter all feature codes are referenced as FC xxxx, where xxxx is the appropriate feature code of the particular function.

Table 5-1 on page 116 identifies the components required to construct a minimum configuration for a p5-590 includes the following items:

Table 5-1 p5-590 minimum system configuration

Quantity	Component description	Feature code (FC)
One	IBM @server p5 590	9119-590
One	Media drawer for installation and service actions (additional media features may be required)	19” 7212-102 or FC 5795
One	16 -way, POWER5 Processor Book, 0-Way Active	FC 7981
Eight	1 -way, Processor Activations	FC 7925
Two	Memory cards with a minimum of 8 GB of activated memory	Refer to the sales guide for valid memory configuration feature codes
Two	Processor Clock Cards, Programmable	FC 7810
One	Power Cable Group, Bulk Power to CEC and Fans	FC 7821
Three	Power Converter Assemblies, Central Electronics Complex	FC 7809
One	Power Cable Group, First Processor Book	FC 7822

Quantity	Component description	Feature code (FC)
Two	System service processors	FC 7811
One	Multiplexer Card	FC 7812
Two	RIO-2 Loop Adapters, Single Loop	FC 7818
One	I/O Drawer Note: requires 4U rack space	FC 5791 or FC 5794
One	Remote I/O (RIO) Cable, 0.6 M Note: used to connect drawer halves	FC 7924
Two	Remote I/O (RIO) Cables, 2.5 M	FC 3168
Two	15,000 RPM Ultra3 SCSI Disk Drive Assemblies	FC 3277 or FC 3278
One	I/O Drawer Attachment Cable Group	FC 6122
One	Slim Line or Acoustic Door Kit	FC 6251 or FC 6252
Two	Bulk Power Regulators	FC 6186
Two	Bulk Power Controller Assemblies	FC 7803
Two	Bulk Power Distribution Assemblies	FC 7837
Two	Line Cords	FC 86xx Refer to your sales guide for specific line cord feature code options
One	Language Specify	FC 9xxx Refer to your sales guide for specific language feature code options

Table 5-2 on page 117 identifies the components required to construct a minimum configuration for a p5-595 includes the following items:

Table 5-2 p5-595 minimum system configuration

Quantity	Component description	Feature code (FC)
One	IBM @server p5 595	9119-595
One	6 -way, POWER5 Processor Book, 0-Way Active	FC 7813 or FC 7988

Quantity	Component description	Feature code (FC)
Note: The following two components must be added to p5-595 servers with one processor book (FC 7813) One - Cooling Group (FC 7807) One - Power Cable Group (FC 7826)		
Sixteen	1 -way, Processor Activations	FC 7815 or FC 7990
Two	Memory cards with a minimum of 8 GB of activated memory	Refer to the sales guide for valid memory configuration feature codes
Two	Processor Clock Cards, Programmable	FC 7810
One	Power Cable Group, Bulk Power to CEC and Fans	FC 7821
Three	Power Converter Assemblies, Central Electronics Complex	FC 7809
One	Power Cable Group, First Processor Book	FC 7822
One	Multiplexer Card	FC 7812
Two	Service processors	FC 7811
Two	RIO-2 Loop Adapter, Single Loop	FC 7818
One	I/O Drawer Note: 4U rack space required	FC 5791 or FC 5794
One	Remote I/O (RIO) Cable, 0.6 M Note: used to connect drawer halves	FC 7924
Two	Remote I/O (RIO) Cables, 3.5 M	FC 3147
Two	15,000 RPM Ultra3 SCSI Disk Drive Assembly	FC 3277 or FC 3278
One	PCI SCSI Adapter or PCI LAN Adapter for attachment of a device to read CD media or attachment to a NIM server	Refer to the sales guide for valid adapter feature code
One	I/O Drawer Attachment Cable Group	FC 6122
One	Slim Line or Acoustic Door Kit	FC 6251 or FC 6252

Quantity	Component description	Feature code (FC)
Two	Bulk Power Regulators	FC 6186
Two	Power Controller Assemblies	FC 7803
Two	Power Distribution Assemblies	FC 7837
Two	Line Cords	FC 86xx Refer to your sales guide for specific line cord feature code options
One	Language Specify	FC 9xxx Refer to your sales guide for specific language feature code options

5.2.2 LPAR considerations

Logical partitioning (LPAR) allows the p5-590 and p5-595 to simultaneously run multiple operating system instances on a single server. LPAR also allows allocation of server resources to various client partitions.

IBM @server p5 systems introduce an enhanced partitioning model, known as *Micro-Partitioning*. This feature allows for finer granularity of POWER5 microprocessor allocation across client partitions. With this enhanced feature, 1/10th of a POWER5 microprocessor can be allocated to a logical partition as a minimum, and 1/100 can be assigned above that. This means that a single POWER5 microprocessor can be shared between ten different logical partitions.

The following list outlines rules when configuring logical partitions in p5-590 and p5-595 servers.

- Logical partition (LPAR) allocation, monitoring, and control is provided by the Hardware Management Console (HMC)
- Each logical partition must support one of the following operating systems:
  - AIX 5L Version 5.2 or Version 5.3
  - i5/OS V5R3, or later, in a partition hosted by AIX 5L (up to 2 CPUs)
  - SUSE LINUX Enterprise Server 9 for POWER systems, or later
  - Red Hat Enterprise Linux AS 3 for POWER, or later
- Each logical partition runs its own instance of a supported operating system
- Non-EEH compatible I/O adapters cannot be used in a partitioned environment.

- Logical partitions must be populated with the following minimum set of system resources
  - One dedicated processor or 0.1 processing unit.
  - 128 MB memory
  - Physical or virtual storage adapter (SCSI card)
  - Physical or virtual network adapter
  - 1 GB storage

Keep in mind the following limitations regarding Micro-Partitioning:

- Shared processor partitions must be configured with at least 0.1 processing units of a physical processor
- The number of shared processor partitions for a system depends on the number of active processors in a system
- p5-590 and p5-595 systems support up to 254 partitions
- Each logical partition supports a maximum of 64 virtual processors
- Mixing dedicated and shared processors within the same partition is not a supported configuration

5.2.3 Processor configuration rules

The p5-590 can be populated with one or two POWER5 standard (1.65 GHz) 16-way processor books. The p5-595 can be populated with up to four POWER5 16-way processor books. Each processor book contains two 8-way multi-chip modules (MCMs). Each 8-way MCM contains four dual core processor chips.

The following list outlines configuration features specific to the MCM processor books on the p5-590 and p5-595.

- Each pair of processors is supported by 1.9 MB of Level 2 (L2) cache and 36 MB of Level 3 (L3) cache
- p5-595 processor books are available two speeds: standard (1.65 GHz) and turbo (1.9 GHz)
- All p5-595 processor books must operate at the same speed
- For both p5-590 and p5-595, each processor book provides sixteen memory card slots and six slots for RIO-2 dual loop adapters for attachment of I/O drawers
- One Multiplexer Card (FC 7812) is required for each processor book to provide a communication channel between the individual processor book and the system service processor

5.2.4 Memory configuration rules

The p5-590 and p5-595 implements a dynamic memory allocation policy that is more sophisticated than the IBM @server pSeries POWER4 systems memory allocation model. Unused physical memory can be assigned to a partition without having to specify the precise address of the assigned physical memory in the partition profile; the system selects the memory resources automatically.

From the hardware perspective the minimum amount of physical memory for each partition is 128 MB. Additional physical memory can be assigned to partitions in increments of 16 MB. At the operating system level, the Virtual Memory Manager (VMM) manages the logical memory within a partition. The POWER hypervisor and the POWER5 microprocessor manage access to the physical memory.

The purpose of this section is to introduce minimum and maximum memory configurations for the p5-590 and p5-595; provide memory installation guidelines; discuss different types of supported memory cards; and provide a summary of overall memory usage and allocation throughout the platform.

In the p5-590 and p5-595, each processor book provides sixteen memory card slots. The p5-590 supports a maximum of 32 memory cards; the p5-595 supports a maximum of 64 memory cards. The p5-590 and p5-595 have the following minimum and maximum configurable memory resource allocation requirements:

- Both p5-590 and p5-595 support a require a minimum of 8 GB of configurable system memory
- p5-590 supports a maximum of 1024 GB of configurable memory
- p5-595 supports a maximum of 2048 GB of configurable memory

> **Note:** 32 GB memory cards to enable maximum memory are planned for availability April 8, 2005.

Use the following guidelines when configuring memory resources within the platform:

- Memory resources must be installed in identical pairs
- p5-590 and p5-595 servers with one processor book must have a minimum of two memory cards installed
- p5-590 and p5-595 servers with two processor books must have a minimum of four memory cards installed per processor book (two per MCM)

Following memory configuration guidelines are recommended:

- The same amount of memory should be used for each MCM (two per processor book) in the system
- No more than two different sizes of memory cards should be used in each processor book
- All MCMs in the system should have the same aggregate memory size
- At least 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

In order to optimize performance for those p5-590 and p5-595 servers being used for high-performance computing, the following are strongly recommended:

- Use DDR2 memory
- Install enough memory to support of each 8-way MCM
- Use the same sized memory cards across all MCMs and processor books in the system

Both the p5-590 and p5-595 utilize dual data rate (DDR) DRAM memory cards. The list below summarizes the types of memory cards available on the p5-590 and p5-595:

- DDR1 and DDR2 memory cards are available on both platform models
- DDR1 and DDR2 memory cards cannot be mixed within a p5-590 or p5-595 server
- The following DDR1 memory is available:
  - 4 GB 266 MHz CUoD card with 2 GB active (FC 7816)
  - 8 GB 266 MHz CUoD card with 4 GB active (FC 7835)
  - 16 GB fully activated 266 MHz card (FC 7828)
  - 32 GB fully activated 200 MHz card (FC 7829)
  - 256 GB package of 32 fully activated 8 GB 266 MHz cards (FC 8195)
  - 512 GB package of 32 fully activated 16 GB 266 MHz cards (FC 8197)
  - 512 GB package of 16 fully activated 32 GB 200 MHz cards (FC 8198)
- DDR2 memory is available in 4 GB, 533 MHz fully activated memory cards (FC 7814)

Memory resources are configured in the p5-590 and p5-595 servers, using the CEC wizard on the Hardware Management Console (HMC). Memory requirements for partitions depends upon partition configuration, I/O resources assigned, and applications used. Memory can be assigned in increments of

16 MB, 32 MB, 64 MB, 128 MB, and 256 MB. The default memory block size varies according to the amount of configurable memory in the system and is summarized in Table 5-3 on page 123.

Table 5-3 Configurable memory-to-default memory block size

Amount of configurable memory	Default memory block size
Less than 4 GB	16 MB
Greater than 4 GB up to 8 GB	32 MB
Greater than 8 GB up to 16 GB	64 MB
Greater than 16 GB up to 32 GB	128 MB
Greater than 32 GB	256 MB

The default memory block size can be changed by using the Logical Memory Block Size option in the Advanced System Management Interface (ASMI). Only a user with administrator authority can change the default memory block size. The managed system must be shut down and restarted in order for the memory block size change to take effect. If the minimum memory amount in any partition profile on the managed system is less than the new default memory block size, the minimum memory amount in the partition profile must change.

Depending on the overall memory in your system and the maximum memory values you choose for each partition, the server firmware must have enough memory to perform logical partition tasks. Each partition has a Hardware Page Table (HPT). The size of the HPT is determined by the maximum memory values established for each partition, and it is based on a ratio of 1/64 of this value.

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

- 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.

When selecting the maximum memory values for each partition, consider the following:

- Maximum values affect the HPT size for each partition
- The logical memory map size for each partition

Note: Page Table is created based on the maximum values defined on partition profile.

The LPAR Validation Tool is useful for determining memory requirements.

5.2.5 Advanced POWER Virtualization

The Advanced POWER Virtualization feature is standard in the p5-590 and p5-595 servers. Advanced POWER Virtualization enables creation of sub-processor partitions that are in units of less than one full processor and enables the physical system I/O to be virtually allocated across partitions. The following configuration rules are specific to Advanced POWER Virtualization.

- One system processors can support up to ten logical partitions
- Using the Advanced POWER Virtualization feature, the p5-590 and p5-595 can be divided into as many as 254 LPARs. System resources can be dedicated to each LPAR
- The Advanced POWER Virtualization feature is standard on the p5-595 and p5-590.
- Advanced POWER Virtualization feature requires AIX 5L Version 5.3 or SUSE LINUX Enterprise Server 9 for POWER

5.2.6 I/O sub-system configuration rules

Chapter 2, “Hardware architecture” on page 17 provides an architectural overview of the p5-590 and p5-595 I/O subsystem. However, there are some configuration rules specific to the I/O sub-system.

Use the following guidelines when attaching I/O drawers to a p5-590 or p5-595 central electronics complex (CEC):

- Each half of the I/O drawer is powered separately
- A maximum of eight I/O drawers can be connected to a p5-590 server
- A maximum of twelve I/O drawers can be connected to a p5-595 server

The p5-590 and p5-595 systems can support the following types of I/O drawers and configurations:

- A minimum of one I/O drawer (FC 5791 or FC 5794) is required per system
  - I/O drawer FC 5791 contains 20 PCI-X slots and 16 disk bays
  - I/O drawer FC 5794 contains 20 PCI-X slots and 8 disk bays
- Existing 7040-61D I/O drawers may be attached to a p5-590 or p5-595 server as additional I/O drawers
  - Only 7040-61D I/O drawers containing FC 6571 PCI-X planars are supported.
  - Any FC 6563 PCI planars must be replaced with FC 6571 PCI-X planars before the I/O drawer can be attached
  - Only adapters supported on the p5-590 or p5-595 feature I/O drawers are supported in 7040-61D I/O drawers, if attached
  - Unsupported adapters must be removed before attaching the drawer to the p5-590 or p5-595 server

The following list identifies a set of guidelines for populating supported I/O drawers on the p5-590 and p5-595:

- One single-wide, blind-swap cassette (equivalent to those in FC 4599) is provided in each PCI or PCI-X slot of the I/O drawer. Cassettes not containing an adapter will be shipped with a *dummy* card installed to help ensure proper environmental characteristics for the drawer. If additional single-wide, blind-swap cassettes are needed, FC 4599 should be ordered.
- Each I/O drawer planar provides ten 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 PCI-X slot supports 32 -bit or 64 -bit adapters
- Integrated Ultra3 SCSI adapters do not support external SCSI device attachments

System I/O drawer connections are always made in loops to help protect against a single point of failure resulting from an open, missing, or disconnected cable. Systems with non-looped configurations could experience degraded performance and serviceability.

System I/O drawers are always connected to the p5-590 or p5-595 CEC using RIO-2 loops operating at 1 GHz. RIO-2 loops connect to the system CEC using RIO-2 Loop Attachment Adapters (FC 7818). Each of these adapters has two ports and can support one RIO-2 loop. Up to six of the adapters can be installed in each 16-way processor book.

I/O drawers may be connected to the CEC in either single-loop or dual-loop mode.

- Single-loop mode connects an entire I/O drawer to the CEC using a single RIO-2 loop. The two I/O planars in the I/O drawer are connected together using a short RIO-2 cable. Single-loop connection requires one loop (2 ports) per I/O drawer.
- Dual-loop mode connects each I/O planar in the drawer to the CEC separately. Each I/O planar is connected to the CEC using a separate RIO-2 loop. Dual-loop connection requires two loops (4 ports) per I/O drawer. Dual-loop mode is recommended whenever possible as it provides the maximum bandwidth between the I/O drawer and the CEC. On initial orders of p5-590 and p5-595 servers, IBM manufacturing will place dual-loop-connected I/O drawers as the lowest numerically designated drawers followed by any single-looped I/O drawers.

Table 5-4 on page 126 and Table 5-5 on page 126 indicate the number of single-looped and double-looped I/O drawers that can be connected to a p5-590 or p5-595 server based on the number of processor books installed.

Table 5-4 p5-590 I/O drawers quantity with different loop mode

	Single-looped	**Dual-looped**
1 Processor book	6	3
2 Processor books	8	6

Table 5-5 p5-595 I/O drawers quantity with different loop mode

	Single-looped	**Dual-looped**
1 Processor book	6	3
2 Processor books	12	6
3 Processor books	12	9
4 Processor books	12	12

5.2.7 Disks, boot devices, and media devices

The p5-590 and p5-595 servers must have access to a device capable of reading CD media or to a network installation manager (NIM) server.

Use the following guidelines when configuring media devices or NIM server:

- FC 5795 provides a rack-based media bay solution for the p5-590 and p5-595 servers.
- External media device models 7212-102, 7210-025, or 7210-030 are devices for reading CD media

- External media devices attach to a PCI SCSI adapter in one of the system I/O drawers
- A NIM server must attach to a PCI LAN adapter in one of the system I/O drawers (an Ethernet adapter is recommended)

A minimum of two internal SCSI hard disks are required per p5-590 or p5-595 server. It is recommended 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
- For AIX5L Version 5.2 and 5.3 logical partitions, consideration should be given to the placement of the AIX rootvg volume group in the first I/O drawer. This allows AIX to boot any time other I/O drawers are found offline during boot.
- If the boot source other than internal disk is configured, the supporting adapter should also be in the first I/O drawer

The p5-590 and p5-595 servers incorporate 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 uninterruptedly power system (UPS) to help ensure against loss of data due to power failures.

5.2.8 PCI and PCI-X slots and adapters

The purpose of this section is to provide configuration guidelines for PCI and PCI-X adapters.

- System maximum limits for adapters and devices may not provide optimal system performance; these limits are given for connectivity and function information.
- 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.
- I/O drawer limitations, in addition to individual adapter limitations, are shown in the feature descriptions section of the sales manual.
- Only EEH adapters are supported in a partitioned server.

5.2.9 Keyboards and displays

The purpose of this section is to provide guidelines for configuring keyboards and displays configuration:

- USB keyboards supporting various language groups can be ordered with the I/O drawer for use with native-attached displays on the p5-590 and p5-595; a three-button USB mouse connects to the USB keyboard
- Various flat panel and CRT displays are available to support graphics adapters

5.2.10 Rack, power, and battery backup configuration rules

The primary system rack is a 24-inch rack with an integrated power subsystem to support the p5-590 and p5-595 system. It provides 42U of rack space and houses the CEC and its components.

Rack configuration

When configuring the system rack, you must consider the following rules:

- The p5-595 server must be installed in a raised floor environment
- All p5-590 and p5-595 racks and Expansion Rack features must have door assemblies installed
- Doors kits containing front and rear doors are available in either slim line or acoustic styles

  Slim line door kit Provides a smaller footprint for use where conservation of space is desired

  Acoustic door kit Provides additional acoustic dampening for use where a quieter environment is desired.
- The height of the system rack or expansion racks (42U) may require special handling when shipping by air or when moving under a low doorway

Power subsystem

The primary system rack always incorporates two bulk power assemblies for redundancy. These provide 350 VDC power for devices located in those racks and associated nonpowered Expansion Racks. These bulk power assemblies are mounted in front and rear positions and occupy the top 8U of the rack. To help provide optimum system availability, these bulk power assemblies should be powered from separate power sources with separate line cords.

The following list summarizes considerations specific to the power sub-system:

- Three DC Power Converters (FC 7809) are always required for miscellaneous CEC components and the first 16-way processor book.
- Three additional DC Power Converters must be added for each additional 16-way processor book.
- The base CEC contains four cooling fans. Cooling Group FC 7807 is required in following condition:
  - p5-590 server with two 16-way processor modules installed.
  - p5-595 with one or more turbo processor books (FC 7813)
  - p5-595 with two or more standard processor books (FC 7988)
- Power cable groups are used to connect the DC Power Converters to the bulk power assembly.
  - FC 7821 provides power for the CEC and four cooling fans
  - FC 7822 provides power for the first processor book
  - FC 7823 provides power for the second processor book
  - FC 7824 provides power for the third processor book of p5-595
  - FC 7825 provides power for the fourth processor book of p5-595
  - FC 7826 provides power for Cooling Group FC 7807

Powered Expansion Rack (FC 5792)

Available in p5-595 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.

Unpowered Expansion Rack (FC 8691)

Available for both p5-590 and p5-595 if additional 24-inch rack space is required. To install the Expansion Rack feature, the side cover of the system rack is removed, the Expansion Rack is bolted to the side, and the side cover is placed on the exposed side of the Expansion Rack. Power for components in the Expansion Rack is provided from the bulk power assemblies in the powered Expansion Rack.

Redundant power

The p5-590 and p5-595 use redundant power throughout its design. It implements redundant bulk power assemblies, bulk power regulators, power controllers, power distribution assemblies, DC Power Converters, and associated cabling. Power for the p5-590 and p5-595 CEC is supplied from DC bulk power assemblies in the system rack. The bulk power is converted to the power levels required for the CEC using DC Power Convertors.

Bulk Power Regulators (FC 6186)

Interface to the bulk power assemblies to help ensure proper power is supplied to the systems components. Bulk Power Regulators are always installed in pairs in the front and rear bulk power assemblies to provide redundancy. The number of bulk power regulators required is configuration-dependent based on the number of processor MCMs and I/O drawers installed.

Redundant Bulk Power Controller (BPC) (FC 7803)

Assemblies are required for the bulk power assemblies. In addition to providing power control, each BPC provides six power connectors for attaching system components.

Redundant Bulk Power Distribution (BPD) Assemblies (FC 7837)

Provide power connections to support the system cooling fan's DC Power Converters contained in the CEC and the I/O drawers. Ten connector locations are provided by each Power Distribution Assembly. Additional Power Distribution Assemblies are added to provide more connections for larger system configurations.

When p5-595 servers contain three or four processor books, the power subsystem in the primary system rack can support only the 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 (FC 5792). In very large configurations, which include battery backup, the unpowered Expansion Rack (FC 8691) is attached to the powered Expansion Rack (FC 5792).

- The power subsystem in the primary system rack is capable of supporting p5-595 servers with one or two processor books installed and up to twelve attached I/O drawers. In such configurations, some I/O drawers are mounted in the unpowered Expansion Rack (FC 8691), which is attached directly to the system rack. The I/O drawers in the FC 8691 rack receive their power from the power subsystem in the system rack.
- p5-595 servers with one or two processor books installed can also use the powered Expansion Rack for mounting any I/O drawers that cannot be mounted in the primary system rack. clients purchasing p5-595 servers with only one or two processor books and more I/O drawers than can be configured in the primary system rack should carefully consider their future system growth requirements to determine which rack combination to purchase.
- The number of Bulk Power Regulators (BPR)(FC 6186) and Bulk Power Distribution (BPD) Assemblies (FC 7837) varies, depending on the number of processor books and I/O drawers and battery backup features(p5-595 only) installed in the p5-590 or p5-595 server. For detail the number of BPRs and

BPDs required for different combination, refer to p5-590 and p5-595 sales manual.

Battery backup

An optional Integrated Battery Backup is available, if desired. The battery backup features are designed to protect against power line disturbances and provide sufficient power to allow an orderly system shutdown in the event that the power sources fail. The battery backup features each require 2U of space in the primary system rack or in the Expansion Rack (FC 8691).

- The primary battery backup (FC 6200) is used to back up the front-mounted power subsystem. It interfaces to the power regulators with the FC 6240 cable. If battery backup is desired, one battery backup (FC 6200) should be ordered for each power regulator in the front power subsystem.
- Redundant battery backup (FC 6201) is used for the rear-mounted power subsystem if redundancy is desired. It interfaces to the power regulators using the FC 6240 cable. If redundant battery backup is desired, one FC 6201 should be ordered for each power regulator in the rear power subsystem.
- The primary battery backup (FC 6200) is a prerequisite to ordering the redundant battery backup (FC 6201).
- If additional external communication and storage devices are required, 7014-T00 or T42 racks should be ordered. There is no limit on the quantity of 7014 racks allowed.

5.2.11 HMC configuration rules

In order to configure and administer a partitions on p5-590 and p5-595 systems, you must attach at least one IBM Hardware Management Console for pSeries (HMC) to the system. It is recomended that a second HMC is configured, and each HMC has a second network connection for service and management. Depending on the partioning-capable pSeries server models, the HMC can be ordered as a feature code or a separate orderable product, as shown in Table 5-6 on page 131. For more information on the HMC, see Chapter 8, “Hardware Management Console overview” on page 195.

Table 5-6 Required Hardware Management Console

Short Product Name	HMC	Note
p5-595	7310-C04 or 7310-CR3	1
p5-590	7310-C04 or 7310-CR3	1
p5-570	7310-C04 or 7310-CR3	2
p5-550	7310-C04 or 7310-CR3	2

Short Product Name	HMC	Note
p5-520	7310-C04 or 7310-CR3	2
p5-510	7310-C04 or 7310-CR3	2
OpenPower 720	7310-C04 or 7310-CR3	2
OpenPower 710	7310-C04 or 7310-CR3	2

1. An HMC is required, and two HMCs are recommended. A private network with the HMC providing DHCP services is mandatory on these systems.
2. An HMC is required if the system is partitioned. The HMC is not required if the system is running as a full system partition.

The Hardware Management Console (HMC) provides a set of functions that are necessary to manage the POWER5 systems when LPAR, Capacity on Demand (CoD) without reboot, inventory and microcode management, and remote power control functions are needed. These functions include the handling of the partition profiles that define the processor, memory, and I/O resources that are allocated to an individual partition.

POWER5 processor-based system HMCs require Ethernet connectivity. Sufficient Ethernet adapters must be available to enable public and private networks if you need both.

The 7310 Model C04 is a desktop model with one native 10/100/1000 Ethernet port, two additional PCI slots for additional Ethernet adapters, two PCI-Express slots, and six USB ports.

The 7310 Model CR3 is a 1U, 19-inch rack-mountable drawer that has two native Ethernet ports, two additional PCI slots for additional Ethernet adapters, and three USB ports.

A zero-priced conversions feature are available for exist IBM 7315-C01/C02/C03/CR2 HMCs from the HMC for POWER4 Licensed Machine Code (FC 0960) to HMC for POWER4 Licensed Machine Code (FC 0961).

Each p5-590 or p5-595 server can be connected to two HMCs for redundancy, if desired. The p5-590 or p5-595 is connected to the HMC through Ethernet connections.

One HMC is capable of controlling multiple pSeries servers. The number of servers each console can control varies by server size and complexity. For details on the number of servers and LPARs supported, visit:

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

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

5.2.12 Cluster 1600 considerations

The p5-590 and p5-595 support the IBM pSeries Cluster 1600 running Cluster Systems Management V1.4 for AIX 5L Version 5.2, Version 5.3, and SUSE LINUX Enterprise Server 9.

5.3 Capacity planning considerations

Capacity planning is a predictive process to determine future computing hardware resources required to support estimated changes in workload. The increased workload on computing resources can be a result of growth in business volumes or the introduction of new applications and functions to enhance the current suite of applications.

The objective of the capacity planning process is to develop an estimate of the system resource required to deliver performance objectives that meet a forecast level of business activity at some specified date in the future. As a result of its predictive nature, capacity planning can only be an approximation at best.

5.3.1 p5-590 and p5-595 general considerations

The p5-590 and p5-595 are at the leading-edge of technology, both in processor power and speed, and in I/O capabilities. To get maximum performance for a client’s system, use the following general recommendations for capacity planning:

Processor configuration

The processor configuration and sizing are critical. After you realize the configuration and sizing, the other components are relatively simple to determine.

Balanced systems

You must size all systems to yield a reasonably balanced system in regard to processor, memory, and disk. This allows for balance from the top down and bottom up:

- The processor or processors must be powerful enough to make efficient use of the memory (caching data for later processing) and to keep the disks busy doing useful work reading information and saving the results.

- The opposite is also true. The disks must be fast enough to feed data into the memory subsystem and the memory to feed the main processor caches fast enough to keep the processors operating at maximum efficiency.

Memory configuration

The optimal amount of memory for a particular system depends upon many factors, not the least being the requirements of the key applications targeted for the system. It is important to note, however, that the size and number of memory cards in the system will determine the maximum system bandwidth that the system can deliver.

To maximize memory performance on p5-590 and p5-595 systems, it is better to install more cards of smaller capacity than fewer cards of larger capacity. Using the same size memory cards across all MCMs and processor books is also recommended.

If the client's performance demands are moderate and memory growth is expected; it is wise to consider using fewer higher density memory cards so that a few memory slots are left open for future growth.

I/O system

Plan for sufficient I/O drawers to support the I/O bandwidths expected and the numbers and types devices to be attached.

The p5-590 and p5-595 architecture allows the bandwidth supported by the system to scale with the number of drawers attached. The total bandwidth required by a system will vary depending on the application.

Once each workload has been characterized, the next step is to identify the numbers and types of I/O devices that will be required to support each workload, and the amount of I/O bandwidth that will be handled by each device. Note that device peak bandwidth may be much higher than sustained bandwidth, and that peak loads rarely occur on all devices simultaneously.

Given the device bandwidths, it is relatively straightforward to determine the total I/O bandwidth required of the system. However, in addition to the maximum bandwidth that I/O drawers can support, there are adapter limits that must be considered. Once the total bandwidth requirements and the total numbers and types of adapters have been determined, the sizing process must match these with the bandwidth and numbers and types of adapters supportable by each I/O drawer.

5.3.2 Further capacity planning considerations

The implementation of the same application in different environments can result in significantly different system requirements. There are many factors that can influence the degree of success in achieving the predicted results. These include changes in application design, the way users interact with the application, and the number of users who may use the applications. The key to successful capacity planning is a thorough understanding of the application implementation and use of performance data collected.

Properly capacity planning of pSeries servers can be difficult since every client environment is unique. Usually there is not enough information to make the best decision. Therefore, you must take a realistic approach. And if you make proper assumptions, you can reach an adequate solution.

For more detailed information about capacity and sizing consideration, refer to publication *IBM @server pSeries Sizing and Capacity Planning: A Practical Guide*, SG24-7071*.*

6

Reliability, availability, and serviceability

The terms reliability, availability, and serviceability (RAS) are widely used throughout the computer industry as an indication of a product’s failure characteristics. RAS refers to a collection of interdependent product attributes that, when taken together, attempt to measure how often a product fails, how quickly it can be repaired, and the overall system disruption caused by a failure.

This chapter describes the various features and mechanisms that are implemented in the p5-590 and p5-595 servers to minimize failures, and isolate and recover them if possible, by providing the following topics:

- Section 6.1, “What’s new in RAS” on page 138
- Section 6.2, “RAS overview” on page 138
- Section 6.3, “Predictive functions” on page 140
- Section 6.4, “Redundancy in components” on page 144
- Section 6.5, “Fault recovery” on page 148
- Section 6.6, “Serviceability features” on page 157
- Section 6.7, “AIX RAS features” on page 166
- Section 6.8, “Linux RAS features” on page 168

6.1 What's new in RAS

From a Reliability, Availability and Serviceability (RAS) standpoint, p5-595 and p5-590 servers extend the heritage of mainframe-proven reliability, availability, and serviceability (RAS) capabilities. They are designed to be more reliable and include features designed to increase availability and to support the new levels of virtualization, building upon the industry leading RAS features delivered in IBMs p5 servers. Powerful new RAS functions include:

- PCI Enhanced Error Handling extended to PCI Processor Host Bridge (PHB).
- Enhanced self-healing capabilities such as ECC (error checking and correction) cache and ECC on the fabric bus.
- New converged redundant service processor design with enhanced function.
- More flexible dynamic CPU deallocation function using Micro-Partitioning technology.

The p5-595 and p5-590 are also backed by worldwide IBM service and support. The one-year end-to-end warranty includes AIX 5L operating system support, hardware fixes, staffed phone hardware support, and call tracking.

6.2 RAS overview

IBMs RAS philosophy employs a well thought out and organized architectural approach to:

- Avoid problems, where possible, with a well engineered design.
- Should a problem occur, attempt to recover or retry the operation.
- Diagnose the problem and reconfigure the system as needed.
- Automatically initiate a repair and call for service.

As a result, IBM servers are recognized around the world for their reliable, robust operation in a wide variety of demanding environments.

Figure 6-1 on page 139 shows the comprehensive IBM RAS philosophy.

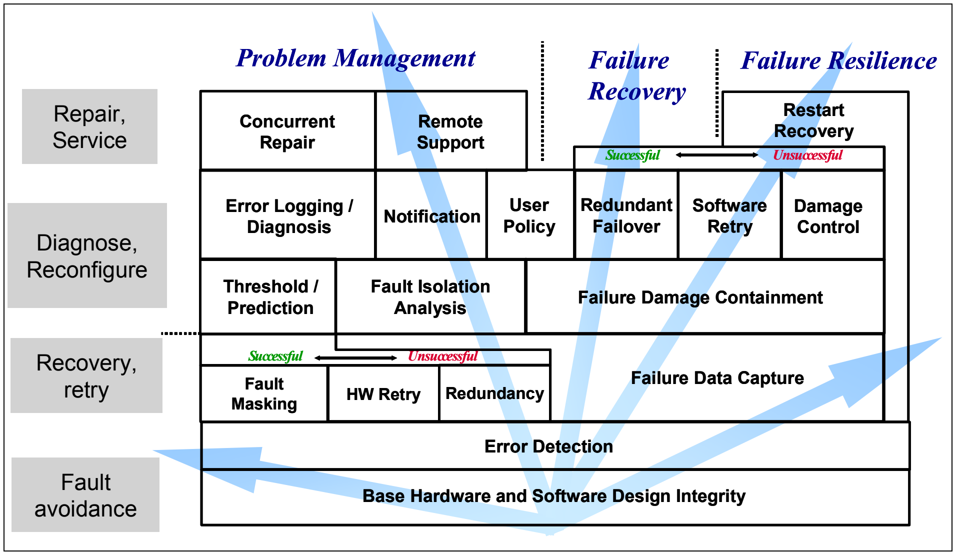


Figure 6-1 IBMs RAS philosophy

Both the p5-595 and p5-590 are designed to provide new levels of proven, mainframe-inspired reliability, availability, and serviceability for mission-critical applications. It comes equipped with multiple resources to identify and help resolve system problems rapidly. During ongoing operation, error checking and correction (ECC) checks data for errors and can correct them in real time. First Failure Data Capture (FFDC) capabilities log both the source and root cause of problems to help prevent the recurrence of intermittent failures that diagnostics cannot reproduce. Meanwhile, Dynamic Processor Deallocation and dynamic deallocation of PCI bus slots help to reallocate resources when an impending failure is detected so applications can continue to run unimpeded.

The p5-595 and p5-590 also include structural elements to help ensure outstanding availability and serviceability. The 24-inch system frame includes hot-swappable disk bays and PCI slots that allow administrators to repair, replace or install components without interrupting the system. Redundant hot-plug power and cooling subsystems provide power and cooling backup in case units fail, and they allow for easy replacement. In the event of a complete power failure, early power off warning capabilities are designed to perform an orderly shutdown. In addition, both primary and redundant battery backup power subsystems are optionally available.

The p5-590 and p5-595 RAS design enhancements can be grouped into four main areas:

Predictive functions These are targeted to monitor the system for possible failures, and take proactive measures to avoid the failures.

Redundancy in components Duplicate components and data paths to prevent single points of failure.

Fault recovery Provide mechanisms to dynamically recover from failures, without system outage. This includes dynamic deallocation of components and hot-swap capability.

Serviceability features Enable the system to automatically call for support, and provide tools for rapid identification of problems.

The p5-590 and p5-595 RAS presents features in all these categories, as described in the following sections.

6.3 Predictive functions

In a mission-critical application, any outage caused by system failure will have an impact on users or processes. The extent of the impact depends on the type of the outage and its duration.

Unexpected system outages are the most critical in a system. The disruption caused by these outages not only interrupts the system execution, but can potentially cause data problems of either loss or integrity. Moreover, the recovery procedures after such outages are generally longer than for planned outages, because they involve error recovery mechanisms and log analysis.

Planned system outages are also critical in a mission-critical environment. However, the impact can be minimized by adequately planning the outage time and the procedures to be executed. Applications are properly shut down, and users are aware of the service stop. Therefore, there is less exposure when doing planned outages in a system.

Reliability engineering is the science of understanding, predicting, and eliminating the sources of system failures. Therefore, the ability to detect imminent failures, and the ability to plan system maintenance in advance are fundamental to reducing outages in a system, and to effectively implement a reliable server.

6.3.1 First failure data capture (FFDC)

Diagnosing problems in a computer is a critical requirement for autonomic computing. The first step to producing a computer that truly has the ability to self heal is to create a highly accurate way to identify and isolate hardware errors. IBM has implemented a server design that builds-in thousands of hardware error check stations that capture and help to identify error conditions within the server. The p5-590 and p5-595 server, for example, includes almost 80,000 checkers to help capture and identify error conditions. These are stored in over 29,000 fault isolation register (FIR) bits. Each of these checkers is viewed as a diagnostic probe into the server, and, when coupled with extensive diagnostic firmware routines, allows quick and accurate assessment of hardware error conditions at run-time.

Named *First Failure Data Capture* (FFDC), this proactive diagnostic strategy is a significant improvement over less accurate reboot and diagnose service approaches. Figure 6-2 illustrates FFDC error checkers and fault isolation registers (FIR). Using projections based on IBM internal tracking information it is possible to predict that high impact outages would occur 2 to 3 times more frequently without a FFDC capability. In fact, without some type of pervasive method for problem diagnosis, even simple problems which occur intermittently can be a cause for serious and prolonged outages.

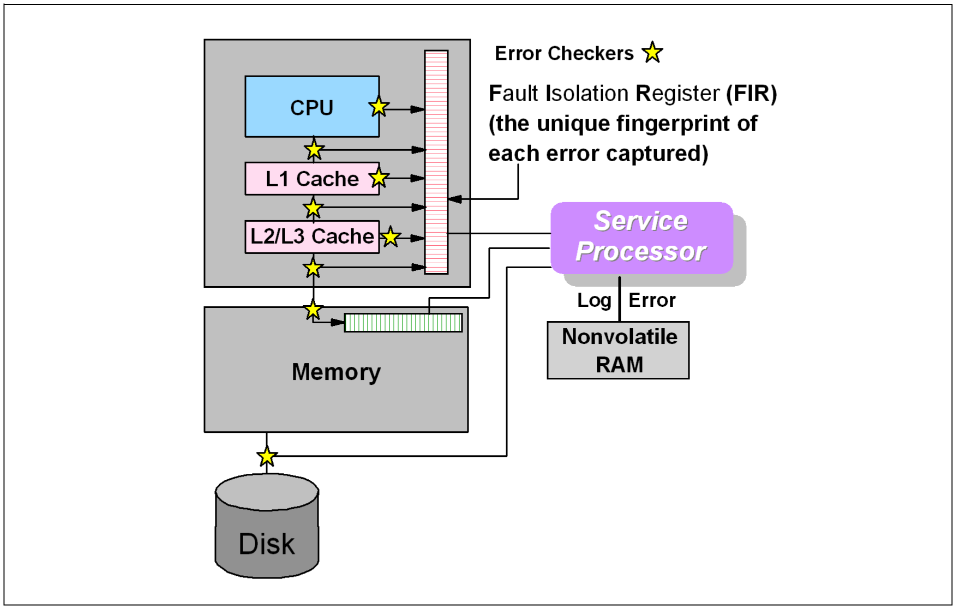


Figure 6-2 FFDC error checkers and fault isolation registers

Integrated hardware error detection and fault isolation has been a key component of IBMs UNIX server design strategy since 1997. FFDC check stations are carefully positioned within the server logic and data paths to ensure that potential errors can be quickly identified and accurately tracked to an individual field replaceable unit (FRU). These checkers are collected in a series of Fault Isolation Registers (FIR), where they can easily be accessed by the service processor. All communication between the service processor and the FIR is accomplished *out of band*. That is, operation of the error detection mechanism is transparent to an operating server. This entire structure is below the architecture and is not seen, nor accessed, by system level activities.

In this environment, strategically placed error checkers are continuously operating to precisely identify error signatures within defined hardware fault domains. IBM servers are designed so that in the unlikely event that a fatal hardware error occurs, FFDC, coupled with extensive error analysis and reporting firmware in the service processor, should allow IBM to isolate a hardware failure to a single FRU. In this event, the FRU part number will be included in the extensive error log information captured by the server. In select cases, a set of FRUs will be identified when the fault is on an interface between two or more FRUs. For example, three FRUs may be called out when the system cannot differentiate between a failed driver on one component, the corresponding receiver on a second, or the interconnect fabric. In either case, it is IBMs maintenance practice for the p5-590 and p5-595 systems to replace all of the identified components as a group. Meeting rigorous goals for fault isolation requires a reliability, availability, and serviceability methodology that carefully instruments the entire system logic design with meticulously placed error checkers.

6.3.2 Predictive failure analysis

Statistically, there are two main situations where a component has a catastrophic failure: Shortly after being manufactured, and when it has reached its useful life period. Between these two regions, the failure rate for a given component is generally low, and normally gradual. A complete failure usually happens after some degradation has happened, be it in the form of temporary errors, degraded performance, or degraded function.

The p5-590 and p5-595 have the ability to monitor critical components such as processors, memory, cache, I/O subsystem, and internal disks, and detect possible indications of failures. By continuously monitoring these components, upon reaching a threshold, the system can isolate and deallocate the failing component without system outage, thereby avoiding a complete failure.

6.3.3 Component reliability

The components used in the CEC provide superior levels of reliability that are available and undergo additional stress testing and screening above and beyond the industry-standard components that are used in several UNIX-based systems today.

Fault avoidance is also enhanced by minimizing the total number of components, and this is inherent in POWER5 chip technology, with two processors per chip. In addition, the basic memory DIMM technology has been significantly improved in reliability through the use of more reliable soldered connections to the memory cards. Going beyond component reliability, an internal array of soft errors throughout the POWER5 chip are systematically masked using internal ECC recovery techniques whenever an error is detected.

The POWER5 chip provides additional enhancements such as virtualization, and improved reliability, availability, and serviceability (RAS) at both chip and system levels. The chip includes approximately 276 M transistors. Given the large number of circuits and the small feature size, one of the biggest challenges in modern processor design is controlling chip power consumption in order to reduce heat creation. Unmanaged, the heat can significantly affect the overall reliable of a server. The introduction of simultaneous multi-threading in POWER5 allows the chip to execute more instructions per cycle per processor core, increasing total switching power. In mitigation, POWER5 chips use a fine-grained, dynamic clock-gating mechanism. This mechanism turns off clocks to a local clock buffer if dynamic management logic determines that a set of latches driven by the buffer will not be used in the next cycle. This allows substantial power saving with no performance impact.

6.3.4 Extended system testing and surveillance

The design of the p5-590 and p5-595 aids in the recognition of intermittent errors that are either corrected dynamically or reported for further isolation and repair. Parity checking on the system bus, cyclic redundancy checking (CRC) on the remote I/O (RIO) bus, and the use of error correcting code on memory and processors contribute to outstanding RAS characteristics.

During the boot sequence, built-in self test (BIST) and power-on self test (POST) routines check the processors, cache, and associated hardware required for a successful system start. These tests run every time the system is powered on.

Additional testing can be selected at boot time to fully verify the system memory and check the chip interconnect wiring. When a system reboots after a hard failure, it performs extended mode tests to verify that everything is working properly and that nothing was compromised by the failure. This behavior can be overridden by the systems administrator.

6.4 Redundancy in components

The p5-590 and p5-595 system design incorporates redundancy in several components to provide fault-tolerant services where failures are not allowed. Power supplies, fans, blowers, boot disks, I/O links, and power cables offer redundancy to eliminate single points of failure. Some of these features are highlighted, as described in the following sections.

6.4.1 Power and cooling redundancy

The p5-590 and p5-595 provide full power and cooling redundancy, with dual power cords and variable-speed fans and blowers, for both the central electronics complex (CEC) and the I/O drawers.

Within the CEC rack, the N+1 power and cooling subsystem provides complete redundancy in case of failures in the bulk or regulated power supplies, the power controllers, and the cooling units, as well as the power distribution cables. As on the zSeries server, concurrent repair is supported on all of the CEC power and cooling components.

There is also a redundant feature called internal battery features (IBF) designed to maintain system operation during short moments of power fluctuation conditions. For full power loss protection, optional uninterruptible power supply (UPS) systems in addition to, or in place of, the IBF features should be used. You should see the IBF feature as a redundancy feature only; it will not replace the UPS capabilities.

In case of a fan or blower failure, the remaining fans automatically increase speed to compensate for the air flow from the failed component.

6.4.2 Memory redundancy mechanisms

There are several levels of memory protection implemented on the p5-590 and p5-595 systems. From the internal L1 caches to the main memory, several features are implemented to assure data integrity and data recovery in case of memory failures.

- Bit steering to redundant memory in the event of a failed memory module to keep the server operational
- Bit-scattering, thus allowing for error correction and continued operation in the presence of a complete chip failure (*Chipkill* recovery)
- Single-bit error correction and double-bit error detection using ECC without reaching error thresholds for main, L2, L3 cache, and fabric bus
- L1 cache is protected by parity and re-fetches data from L2 cache when errors are detected
- L3 cache line deletes extended from 2 to 10 for additional self-healing
- Memory scrubbing to help prevent soft-error memory faults

Figure 6-3 graphically represents the redundancy and error recovery mechanisms on the main memory.

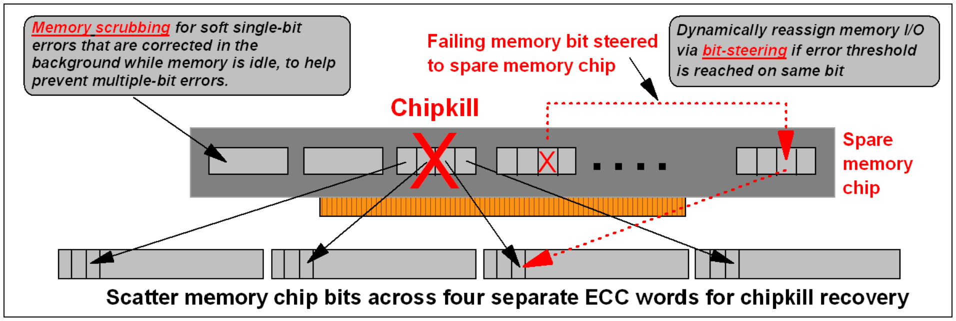

Memory scrubbing for soft single-bit errors that are corrected in the background while memory is idle, to help prevent multiple-bit errors.
Failing memory bit steered to spare memory chip
Dynamically reassign memory I/O via bit-steering if error threshold is reached on same bit
Chipkill
Spare memory chip
Scatter memory chip bits across four separate ECC words for chipkill recovery

Figure 6-3 Memory error recovery mechanisms

Uncorrectable error handling

While it is a rare occurrence, an uncorrectable data error can occur in memory or a cache, despite all precautions built into the server. In servers prior to IBMs POWER4 processor-based offerings, this type of error would eventually result in a system crash. The IBM eServer p5 systems extend the POWER4 technology design and include techniques for handling these errors.

On these servers, when an uncorrectable error (UE) is identified at one of the many checkers strategically deployed throughout the system's central electronic complex, the detecting hardware modifies the ECC word associated with the data, creating a special ECC code. This code indicates that an uncorrectable error has been identified at the data source and that the data in the standard ECC word is no longer valid. The check hardware also signals the service processor and identifies the source of the error. The service processor then takes appropriate action to handle the error. This technique is named special uncorrectable error (SUE) handling.

No system or partition activity is terminated based on simple error detection. In many cases, a UE will cause generation of a synchronous machine check interrupt. The machine check interrupt occurs when a processor tries to load the incorrect data. The firmware provides a pointer to the instruction that referred to the incorrect data, the system continues to operate normally, and the hardware observes the use of the data.

The system is designed to mitigate the problem using a number of approaches:

- If, as may sometimes be the case, the incorrect data is never actually used, but is simply over-written, then the error condition can safely be voided and the system will continue to operate normally.
- If the incorrect data is actually referenced for use by a process (in AIX 5L Version 5.2 or greater) then the OS is informed of the error. The operating system (OS) will terminate only the specific user process associated with the corrupt data.
- If the incorrect data were destined for an I/O subsystem managed by the AIX kernel, or for the kernel itself, then only the partition associated with the data would be rebooted. All other system partitions would continue normal operation.
- If the incorrect data is written to disk, the I/O hardware detects the presence of SUE and the I/O transaction is terminated.
- Finally, only in the case where the incorrect data is used in a critical area of the POWER Hypervisor would the entire system be terminated and automatically rebooted, preserving overall system integrity.

L3 cache protection

The L3 cache is protected by ECC and SUE handling. The L3 cache also incorporates technology to handle memory cell errors using a special cache line delete algorithm.

During CEC initial program load (IPL), if a solid error is detected during L3 initialization, a full L3 cache line will be deleted. During system runtime, a correctable error is reported as a recoverable error to the service processor. If an individual cache line reaches its predictive error threshold, it will be dynamically deleted. The state of L3 cache line delete will be maintained in a deallocation record and will persist through system IPL. This ensures that cache lines varied offline by the server will remain offline should the server be rebooted. These error prone lines can not then cause system operational problems. In the IBM @server p5 product family, the server can dynamically delete up to 10 cache lines. It is not likely that deletion of a couple of cache lines will adversely affect server performance. If this total is reached, the L3 is marked for persistent deconfiguration on subsequent system reboots until repair.

Array recovery and array persistent deallocation

Array persistent deallocation refers to the fault resilience of the arrays in a POWER5 microprocessor. The L1 I-cache, L1 D-cache, L2 cache, L2 directory and L3 directory all contain redundant array bits. If a fault is detected, these arrays can be repaired during IPL by replacing the faulty array bit(s) with the built-in redundancy, in many cases avoiding a part replacement.

The initial state of the array repair data is stored in the FRU vital product data (VPD) by manufacturing. During the first server IPL, the array repair data from the VPD is used for initialization. If an array fault is detected in an array with redundancy by the array built-in-self-test diagnostic, the faulty array bit is replaced. Then the updated array repair data is stored in the service processor persistent storage as part of the deallocation record of the processor. This repair data is used for subsequent system boots.

During system run time, the service processor monitors recoverable errors in these arrays. If a predefined error threshold for a specific array is reached, the service processor tags the error as pending in the deallocation record to indicate that the error is repairable by the system during next system IPL. The error is logged as a predictive error, repairable using re-IPL, avoiding a FRU replacement if the repair is successful.

A further refinement is included in the design of the error handling logic for POWER5 processors. In this processor, an L2 cache can be considered as being built of three slices, three semi-independent sections. If an error that can not be repaired using the technique described above occurs in the L2 directory or in one of the slices, the system can operate on either the first or third cache slice. The firmware will use array persistent deallocation to turn off system access to the remaining two cache slices on the next server boot. As a result, the system will boot with the two processors active but sharing only 1/3 of the normally available L2 cache. An error message will indicate the need for deferred repair of the processor FRU.

For all processor caches, if repair on reboot does not fix the problem, the processor containing the cache can be deconfigured.

6.4.3 Service processor and clocks

A number of availability improvements have been included in the service processor in the p5-590 and p5-595 servers. Code access is CRC protected. The POWER Hypervisor firmware performs the initialization and configuration of the POWER5 processor, as well as the virtualization support required to run up to 254 partitions concurrently on the p5-590 and p5-595 server. The POWER Hypervisor supports many advanced functions, including processor sharing, virtual I/O, high-speed communications between partitions using Virtual LAN and concurrent maintenance. Maintaining two copies ensures that the service processor can run even if a Flash memory copy becomes corrupted, and allows for redundancy in the event of a problem during the upgrade of the firmware. In addition, if the service processor encounters an error during run-time, it can reboot itself while the server system stays up and running. There will be no server application impact for service processor transient errors. If the service processor encounters a code hang condition, the POWER Hypervisor can detect the error and direct the service processor to reboot, avoiding other outages. Depending on configuration, a system with multiple building blocks (IBM @server p5 595, p5 590, or p5 570 servers) may be ordered with two service processors and two system clocks.

6.4.4 Multiple data paths

The I/O subsystem on the p5-590 and the p5-595 is based on the Remote I/O link technology. This link uses a loop interconnect technology to provide redundant paths to I/O drawers. Each I/O drawer is connected to two RIO ports, and each port can access every component in the I/O drawer. During normal operations the I/O is balanced across the two ports. If a RIO link fails, the hardware is designed to automatically initiate a RIO bus reassignment to route the data through the alternate path to its intended destination. Any break in the loop is recoverable using alternate routing through the other link path and can be reported to the service provider for a deferred repair.

6.5 Fault recovery

The p5-590 and p5-595 offer new features to recover from several types of failures automatically, without requiring a system reboot. The ability to isolate and deconfigure components while the system is running is of special importance in a partitioned environment, where a global failure can impact different applications running on the same system.

Some faults require special handling. We will discuss these in the following sections.

6.5.1 PCI bus error recovery

PCI bus errors, such as data or address parity errors and time outs, can occur during either a Direct Memory Access (DMA) operation being controlled by a PCI device, or on a load or store operation being controlled by the host processor.

During DMA, a data parity error results in the operation being aborted, which usually results in the device raising an interrupt to the device driver, allowing the driver to attempt recovery of the operation.

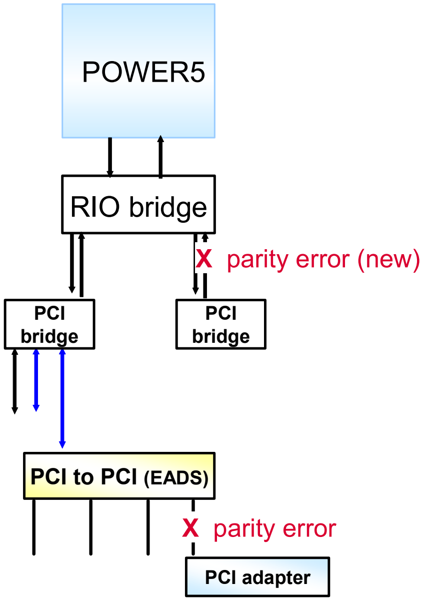
However, all above error scenarios are difficult to handle as orderly. On previous systems, these errors resulted in a bus critical error, followed by a machine check interrupt and system termination.

IBM introduced a methodology in the POWER4 processor-based servers that the I/O drawer hardware, system firmware, and AIX interaction has been designed to allow transparent recovery of intermittent PCI bus parity errors, and orderly transition to the I/O device unavailable state in the case of a permanent parity error in the PCI bus. This mechanism is known as the PCI Extended Error Handling (EEH).

Standard server PCI implementations can detect errors on the PCI bus, but cannot limit the scope of the damage, so system termination is often the result. EEH allowing each PCI slot to have its own PCI bus. Each adapter can therefore be isolated in the case of an error. This enables error recovery to occur without affecting any other adapters on the system. Without EEH, pSeries and p5 machines would checkstop in the event of a PCI bus error, either caused by the bus or a device on the bus. EEH brings the function to freeze an adapter in the event of an I/O error and avoid the checkstop. An adapter reset is tried and is allowed to fail three times before the adapter is marked as dead.

Use of EEH is particularly important on the p5-590 and p5-595 running with multiple partitions. If non-EEH enabled PCI adapters are in use, it is possible for multiple partitions to be disrupted by PCI bus errors. If a client's application requires the use of non-EEH enabled PCI adapters, careful placement of those PCI adapters in the I/O drawer can limit PCI bus error disruptions to a single logical partition. It is for this reason that any non-EEH capable adaters will be disabled when a p5 server is partitioned.

POWER5 processor-based servers extend the capabilities of the EEH methodology as shown in Figure 6-4 on page 150. Generally, on POWER5 platforms, all PCI adapters controlled by operating system device drivers are connected to a PCI secondary bus created through an IBM designed PCI-PCI bridge. This bridge isolates the PCI adapters and supports hot plug by allowing program control of the power state of the I/O slot. PCI bus errors related to individual PCI adapters under partition control can be transformed into a PCI slot freeze condition and reported to the EEH device driver for error handling. Errors that occur on the interface between the PCI-PCI bridge chip and the Processor Host Bridge (the link between the processor remote I/O bus and the primary PCI bus) result in a *bridge freeze* condition, effectively stopping all of the PCI adapters attached to the bridge chip. An operating system may recover an adapter from a bridge freeze condition by using POWER Hypervisor functions to remove the bridge from freeze state and resetting or reinitializing the adapters.



The IBM POWER5 systems add additional recovery features to handle potential errors in the Processor Host Bridge (PCI bridge), and the GX+ bus adapter. These new servers also support “hot” add and removal of entire I/O drawers. These features provide improved diagnosis, isolation, and management of errors in the server I/O path and new opportunities for concurrent maintenance - to allow faster recovery from I/O path errors, often without impact to system operation.

Figure 6-4 EEH on POWER5

The ultimate situation is to use only EEH-enabled PCI adapters, to eliminate system and partition disruptions due to PCI bus errors, and merely suffer the loss of a single PCI adapter if that adapter causes a PCI bus error. Most adapters support EEH. EEH is part of the PCI 2.0 specification, although its implementation is not required for PCI compliance.

For those older adapters that do not support EEH, special care should be taken and documentation should be checked to ensure they are currently supported.

6.5.2 Dynamic CPU deallocation

Dynamic CPU deallocation has been available since AIX Version 4.3.3 on previous RS/6000 and pSeries systems, is the ability for a system to automatically deconfigure an error prone CPU before it causes an unrecoverable system error (unscheduled server outage). It is part of the p5-590 and p5-595 RAS features.

CPU dynamic deconfiguration relies on the service processor's ability to use FFDC generated recoverable-error information and to notify the AIX operating system when the CPU reaches its predefined error limit. AIX will then *drain* the run-queue for that CPU, redistribute the work to the remaining CPUs, deallocate the offending CPU, and continue normal operation, although potentially at a lower level of system performance. While AIX Version 4.3.3 precluded the ability for a SMP server to revert to a uniprocessor (for example, a 2-way to a 1-way configuration), this limitation was lifted with AIX 5L Version 5.1.

AIX 5L Version 5.2 support for dynamic logical partitioning (DLPAR) allowed additional system availability improvements. An IBM @server p5 server that includes an unlicensed CPU (an unused CPU included in a Capacity Upgrade on Demand (CUOD) system configuration) can be configured for CPU hot sparing. In this case, as a system option, the unlicensed CPU can automatically be used to *back-fill* for the deallocated bad processor. In most cases, this operation is transparent to the system administrator and to end users. The spare CPU is logically moved to the target system partition, AIX moves the workload, and the failing processor is deallocated. The server continues normal operation with full function and full performance. The system will generate an error message for inclusion in the error logs calling for deferred maintenance of the faulty component.

Refer to Chapter 4, “Capacity on Demand” on page 83 for more detail about CoD.

POWER5 technology and AIX 5L Version 5.3 introduce new levels of virtualization, supporting Micro-Partitioning technology, allowing individual processors to run as many as ten copies of the operating system. These new capabilities allow improvements in the CPU hot spare strategy. POWER5 chips will support both dedicated processor logical partitions and shared processor dynamic LPAR. Dedicated processor partitions, supporting AIX Version 5.2 and V5.3, operate like POWER4 processor-based system logical partitions. In a dedicated processor LPAR, one or more physical CPUs are assigned to the partition.

In shared processor partitions, supported by AIX 5L Version 5.3, a shared pool of physical processors is defined. This shared pool consists of one or more physical processors. In this environment, partitions are defined to include virtual processor and processor entitlements. Entitlements can be considered to be performance equivalents.

In dedicated POWER5 processor partitions, CPU sparing is transparent to the operating system. When a CPU reaches its error threshold, the service processor notifies the POWER Hypervisor to initiate a deallocation event.

- If a CoD processor is available, the POWER Hypervisor automatically substitutes it for the faulty processor and then deallocates the failing CPU.
- If no CoD processor is available, the POWER Hypervisor checks for excess processor capacity (capacity available because processors are unallocated or because one or more partitions in the shared pool are powered off). The POWER Hypervisor substitutes an available processor for the failing CPU.
- If there are no available processors, the operating system is asked to deallocate the CPU. When the operating system finishes the operation, the POWER Hypervisor stops the failing CPU.

In shared processor partitions, CPU sparing operates in a similar fashion as in dedicated processor partitions.

- In this environment, the POWER Hypervisor is notified by the service processor of the error. As previously described, the system first uses any CoD processors.
- Next, the POWER Hypervisor determines if there is at least 1.00 processor unit’s worth of performance capacity available, and if so, stops the failing processor and redistributes the workload.
- If the requisite spare capacity is not available, the POWER Hypervisor will determine how many processor capacity units each partition will need to relinquish to create at least 1.00 processor capacity units. The POWER Hypervisor uses an algorithm based on partition utilization and the defined partition minimum and maximums for CPU equivalents to calculate capacity units to be requested from each partition. The POWER Hypervisor will then notify the operating system (via an error entry) that processor units and/or virtual processors need to be varied off. Once a full processor equivalent is attained, the CPU deallocation event occurs.
- The deallocation event will not be successful if the POWER Hypervisor and OS cannot create a full processor equivalent. This will result in an error message and the requirement for a system administrator to take corrective action. In all cases, a log entry will be made for each partition that could use the physical processor in question.

6.5.3 CPU Guard

It is necessary that periodic diagnostics not run against a processor already found to have an error by a current error log entry. CPU Guard provides the required blocking to prevent the multiple logging of the same error.

6.5.4 Hot-swappable components

Both the p5-590 and p5-595 provide many parts as hot-swappable Field Replaceable Units (FRUs). This feature allows you to replace most p5-590 and p5-595 components concurrently without the need to power off the system. There are parts such as MCMs. L3 cache, memory, and so on, that still require a scheduled maintenance window to perform the replacement. Table 6-1 provides an overview of what components are hot-plug.

Table 6-1 Hot-swappable FRUs

Processor subsystem FRUs	Hot-swappable concurrent maintenance
Blowers	Yes
DCA	Yes
Bulk power enclosure	Yes
Bulk power controller (BPC)	Yes
Bulk power regulator (BPR)	Yes
Bulk power distributor (BPD)	Yes
Bulk power fan (BPF)	Yes
UEPO switch panel	Yes
Internal battery feature (IBF)	Yes
Capacitor book	No
Processor subsystem chassis	No
MCM	No
Memory cards	No
L3 cache	No
Clock card	No
I/O books	No
I/O subsystem FRUs	**Hot-swappable concurrent maintenance**
I/O backplane and riser card	No
DASD backplane	No
Disk drives	Yes
DCA (power supplies)	Yes

Processor subsystem FRUs	Hot-swappable concurrent maintenance
I/O fan assemblies	No
PCI adapters	Yes

6.5.5 Hot-swappable boot disks

The I/O drawer provides up to 16 hot-swappable bays for internal disks, organized in four cages, each one connected to a separate Ultra3 SCSI controller.

Disk mirroring is strongly suggested for the operating system disks in order to eliminate system outages because of operating system disk failures. If a mirrored disk fails, AIX automatically reroutes the I/O requests to the other disk, without service interruption. You can replace the failed drive with the system online, and mirror again, to restore disk redundancy. Mirroring rules are operating system specific and therefore type and architecture of OS should be considered when calculating the number of redundant disks.

6.5.6 Blind-swap PCI adapters

All PCI slots are PCI 2.2-compliant and are hot-plug enabled, which allows most PCI adapters to be removed, added, or replaced without powering down the system. This function enhances system availability and serviceability.

IBM introduced blind-swap mechanism to provide concurrent adding or removal of PCI adapters when the system is running.

Figure 6-5 shows a basic drawing of the blind-swap cassette.

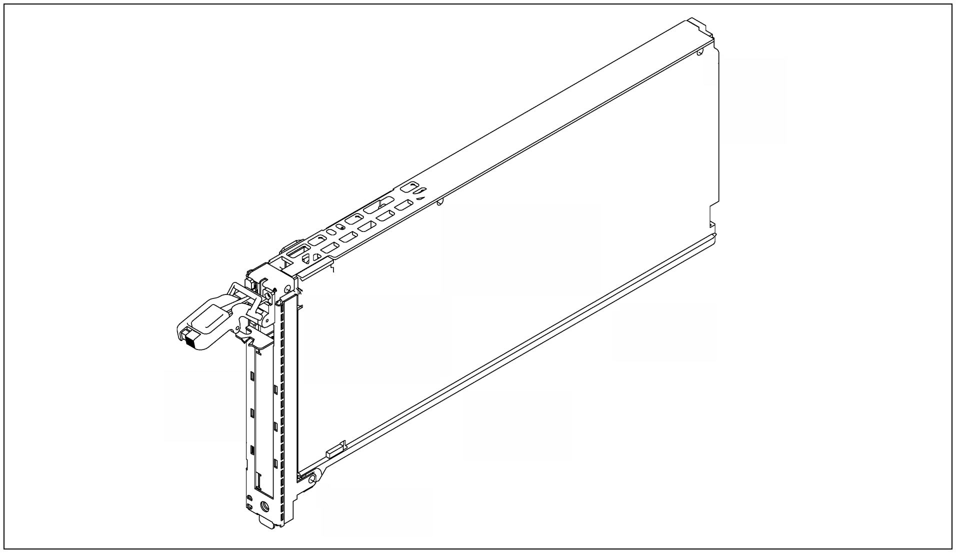


Figure 6-5 Blind-swap cassette

Blind-swap adapters mount PCI I/O cards in a carrier that can be slid into the rear of a server or I/O drawer. The carrier is designed so that the card is guided into place on a set of rails and seated in the slot, completing the electrical connection, by simply shifting an attached lever. This capability allows the PCI adapters to be concurrently replaced without having to put the I/O drawer into a service position. Since first delivered, minor carrier design adjustments have improved an already well-thought out service design. This technology has been incorporated in selected IBM @server p5 servers and I/O drawers.

The hot-plug LEDs outside the I/O drawer indicate whether an adapter can be plugged into or removed from the system. The hot-plug PCI adapters are secured with retainer clips on top of the slots; therefore, you do not need a screwdriver to add or remove a card, and there is no screw that can be dropped inside the drawer. Just in case of exchanging an PCI adapter from the blind-swap cassette, a screwdriver is needed to remove or replace it from the cassette.

The function of hot-plug is not only provided by the PCI slot, but also by the function of the adapter. Most adapters are hot-plug, but some are not. Be aware that some adapters must not be removed when the system is running, such as the adapter with the operating system disks connected to it or the adapter that provides the system console.

6.5.7 Guiding Light Diagnostics

The p5-590 and p5-595 use a technology named Guiding Light Diagnostics, which provides a visual identification of a failed component in order to facilitate the maintenance. This function is available for detecting problems in I/O drawers, PCI planers, fans, blowers, and disks. Guiding Light diagnostics use a series of LEDs (Light Emitting Diodes) to quickly guide a client or client engineer to a failed hardware component so that it can be repaired or replaced.

In the Guiding Light LED implementation, when a fault condition is detected on the p5-590 and p5-595 system, an amber System Attention LED will be illuminated. Upon arrival at the server, a CE sets the identify mode, selecting a specific problem to be identified for repair by the Guiding Light method. The Guiding Light system pinpoints the component by flashing the amber identity LED associated with the part to be replaced.

The system not only clearly identify components for replacement by using specific component level indicators, but can also guide the service representative directly to the component by signaling (causing to flash) the Rack/Frame System Identify indicator and the Drawer Identify indicator on the drawer containing the component. The flashing identify LEDs direct the service representative to the correct system, the correct enclosure, and the correct component.

Disk drives, fans, blowers, and PCI adapters have specific LEDs to indicate the operational state and faults when they occur. Figure 6-6 on page 156 shows the LED placement on I/O drawers from the front view.

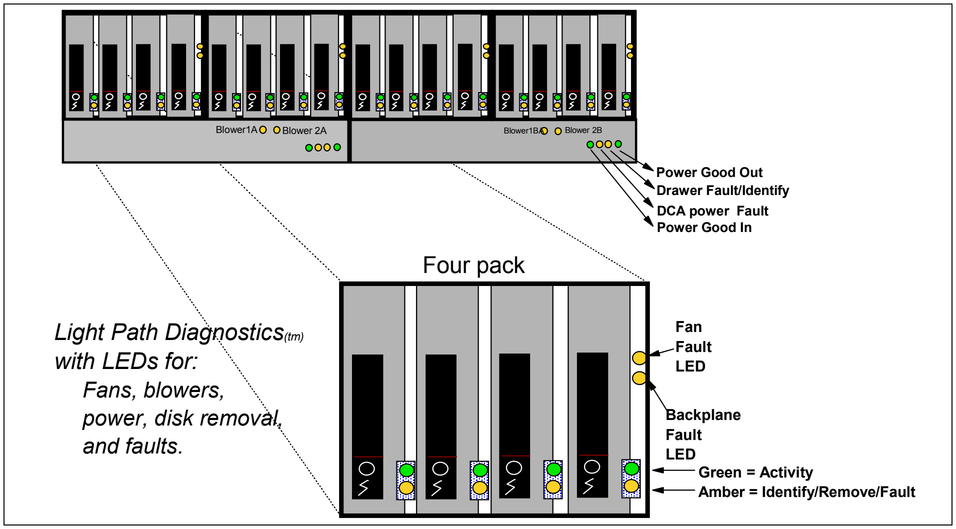


Figure 6-6 Status LEDs for components in I/O drawers

6.6 Serviceability features

The service strategy for the IBM @server p5 families evolves from, and improves upon, the service architecture deployed on IBM pSeries servers. The service team has enhanced the base service capability and continues to implement a strategy that incorporates best-of-breed service characteristics from various IBM @server systems including the xSeries, iSeries, pSeries, and zSeries systems.

The service goal is to provide the most efficient service environment by designing a system package that incorporates:

- Easy access to service components
- On demand service education
- An automated or guided repair strategy using common service interfaces for a converged service approach across multiple IBM server platforms.

The aim is to deliver faster and more accurate repair while reducing the possibility for human error.

6.6.1 Converged service architecture

The IBM @server p5 and @server i5 systems represent a significant convergence of platform service architectures, merging the best characteristics of the iSeries and pSeries product offerings. This union allows similar maintenance approaches and common service user interfaces. A service representative can be trained on the maintenance of the base hardware platform, service tools, and associated service interface and be proficient in problem determination and repair for either POWER5 processor-based platform offering. In some cases, additional training may be required to allow support of I/O drawers, adapters, and devices.

The convergence plan incorporates critical service topics.

- Identifying the failing component through architected error codes.
- Pinpointing the faulty part for service using location codes and LEDs as part of the Guiding Light Diagnostic strategy.
- Ascertaining part numbers to quickly and efficiently order replacement components.
- Collecting system configuration information using common Vital Product Data which completely describes components in the system, to include detailed information such as their point of manufacture and Engineering Change level.

- Enabling service applications, such as Firmware and Hardware EC Management (described below) and Service Agent, to be portable across the multiple hardware and operating system environments.

The resulting commonality makes possible reduced maintenance costs and lower total cost of ownership for IBM @server p5 and @server i5 systems. This core architecture provides consistent service interfaces and a common approach to service, enabling owners of selected @server p5 or @server i5 servers to successfully perform set-up, manage and carry out maintenance, and install server upgrades; all at their own schedule and without available IBM support personnel.

6.6.2 Hardware Management Console

The Hardware Management Console (HMC) is an independent workstation used by system administrators to setup, manage, configure, and boot their IBM @server p5 or @server i5 server. The HMC for @server p5 system includes improved performance, enabling system administrators to define and manage sub-processor partitioning capabilities and virtual I/O features; advanced connectivity; and sophisticated firmware performing a wide variety of systems management and service functions.

One significant improvement on the IBM @server p5 and @server i5 system HMC is to replace the serial attachment method used on predecessor consoles, with a LAN interface allowing high bandwidth connections to servers.

With the exception of the p5-590 and p5-595, administrators can choose to establish a private service network, connecting all of their POWER5 processor-based servers and management consoles. Or they can include their service connections in their standard operations network. The Ethernet LAN interface also allows the HMC to be placed physically farther away from managed servers, though for service purposes it is still desirable to install the HMC in close proximity to the systems (within 8 meters or 25 feet is recommended).

The HMC comes with an install wizard to assist with installation and configuration of the HMC itself. This wizard helps to reduce user errors by guiding administrators through the configuration steps required to successfully install the HMC operating environment.

The Hardware Management Console provides a number of RAS features to the servers it manages.

- Automated Install/Maintenance/Upgrade

The HMC provides a variety of automated maintenance procedures to assist in problem determination and repair. The Hardware Management Console extends this innovative technology, providing automated install and automated upgrade assistance. These procedures are expected to reduce or eliminate service representative induced failures during the install or upgrade processes.

- Concurrent Maintenance and Upgrade

  All IBM POWER5 processor-based servers provide at least the same level of concurrent maintenance capability as was available in their predecessor pSeries (POWER4) servers. Components such as power supplies, fans, blowers, disks, HMCs, PCI adapters and devices can be repaired concurrently (hot service and replace).

6.6.3 Error analyzing

Since 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 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 off-line (even if a redundant component has been used for sparing).

When a recoverable and serviceable error (the third type above) 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 will send an error notification to a client designated user (by default, the root user). This action also invokes the Service Agent application which initiates appropriate system serviceability actions.

- On servers that do not include an HMC, the 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.

- On servers equipped with an HMC, 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.

In either case, failure information including:

- The source of 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 along probable replacement hardware components when a service call is placed, minimizing system repair time.

In a multi-system configuration, any HMC-attached IBM @server p5 or @server i5 server can be configured to forward call home requests to a central Service Agent Gateway (SAG) application on an HMC which owns a modem and performs the call home on behalf of any of the servers.

Figure 6-7 on page 161 shows the error reporting structure of POWER5.

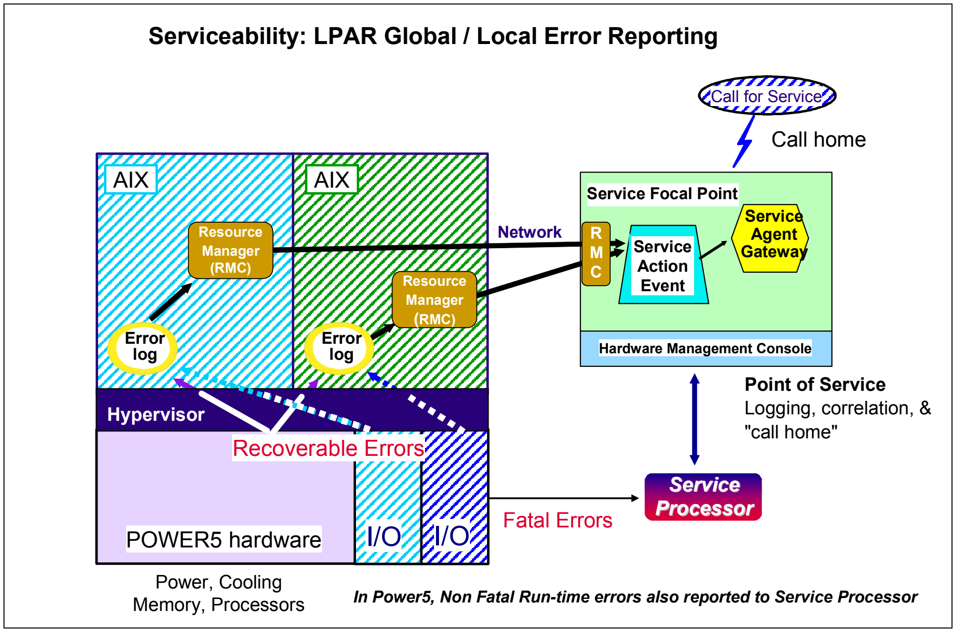


Figure 6-7 Error reporting structure of POWER5

6.6.4 Service processor

The service processor provides for excellent RAS service features such as first failure data capture analysis explained in the prior availability section and surveillance monitoring described previously. It also provides functions such as; power-on and off of the system, reading the service processor and POST error logs, reading vital product data (VPD), changing the bootlist, viewing boot sequence history, and changing service processor configuration parameters, all of which can be performed remotely. clients can enable console mirroring on the system console so they can monitor all remote console activity. For this option to work, a modem must be attached to one of the serial ports and configured appropriately.

The service processor in the p5-590 and p5-595 product is an improved design when compared to the service processor that was available in the IBM POWER4 processor-based systems. The main service processor function in p5-590 and p5-595 is located in the CEC and is based on the new hardware. The CEC contains two service processor cards. The new SP subsystem is UNIX-based and directly drives Ethernet ports to connect an external HMC. The service processor provides services as following:

- Environmental monitoring

The service processor monitors the server's built-in temperature sensors, sending instructions to the system fans to increase rotational speed when the ambient temperature is above the normal operating range.

Using an architected operating system call, the service processor notifies the operating system of potential environmental related-problems (for example, air conditioning and air circulation around the system) so that the system administrator can take appropriate corrective actions before a critical failure threshold is reached.

The service processor can also post a warning and initiate an orderly system shutdown for a variety of other conditions:

- When the operating temperature exceeds the critical level.
- When the system fan speed is out of operational specification.
- When the server input voltages are out of operational specification.

- Mutual Surveillance

  The service processor monitors the operation of the POWER Hypervisor firmware during the boot process and watches for loss of control during system operation. It also allows the POWER Hypervisor to monitor service processor activity. The service processor can take appropriate action, including calling for service, when it detects the POWER Hypervisor firmware or the operating system has lost control. Likewise, the POWER Hypervisor can request a service processor repair action if necessary.

- Availability

  The auto-restart (reboot) option, when enabled, can reboot the system automatically following an unrecoverable software error, software hang, hardware failure, or environmentally induced (AC power) failure.

- Fault Monitoring

  BIST (built-in self-test) and POST (power-on self-test) check processor, L3 cache, memory, and associated hardware required for proper booting of the operating system, when the system is powered on at the initial install or after a hardware configuration change (an MES upgrade). If a non-critical error is detected or if the error occurs in a resource that can be removed from the system configuration, the booting process is designed to proceed to completion. The errors are logged in the system nonvolatile random access memory (NVRAM). When the operating system completes booting, the information is passed from the NVRAM into the system error log where it is analyzed by error log analysis (ELA) routines. Appropriate actions are taken to report the boot time error for subsequent service if required.

  Disk drive fault tracking can alert the system administrator of an impending disk failure before it impacts client operation.

The AIX or Linux log (where hardware and software failures are recorded) is analyzed by ELA routines which warn the system administrator about the causes of system problems.

- Others

  The service processor also manages the interfaces for connecting Uninterruptible Power Source (UPS) systems to the IBM POWER5 processor-based systems, performing Timed Power-On (TPO) sequences, and interfacing with the power and cooling subsystem.

The new service processor incorporates enhanced hardware functions such as an Ethernet service interface, additional serial port communications ports, and larger storage capacity. All of these new features support improved functions for service.

One important service processor improvement gives the system administrator or service representative the ability to dynamically access the Advanced Systems Management Interface (ASMI) menus. In previous generations of servers these menus were only accessible when the system was in standby power mode. Now the menus are now available from any Web browser-enabled console attached to the Ethernet service network concurrent with normal system operation. A user with the proper access authority and credentials can now dynamically modify service defaults, interrogate service processor progress and error logs, set and reset Guiding Light LEDs, indeed, access all service processor functions without having to power-down the system to the standby state.

Separate copies of service processor microcode are stored in separate Flash memory storage areas. Code access is CRC protected. SP resets can be initiated by either the SP itself or the POWER Hypervisor. If the service processor encounters an error during run-time, it can reboot itself while the server system stays up and running. There will be no server application impact for service processor transient errors. If the service processor encounters a code hang condition, the POWER Hypervisor can detect the error and direct the service processor to reboot, avoiding other outage. A service processor reset or reload is not disruptive and will not impact system operation. In each case, the system, if necessary, will initiate a smart dump of the SP control store to assist with problem determination if required.

6.6.5 Service Agent

Electronic Service Agent (also known as Service Agent) is an application program that runs on either AIX or Linux to monitor the system for hardware errors. On pSeries systems managed by the HMC, the primary path for system hardware errors detection and analysis consists of the diagnostics function provided by AIX, the service processor, and the Service Focal Point. Service Agent provides the transport facility to IBM.

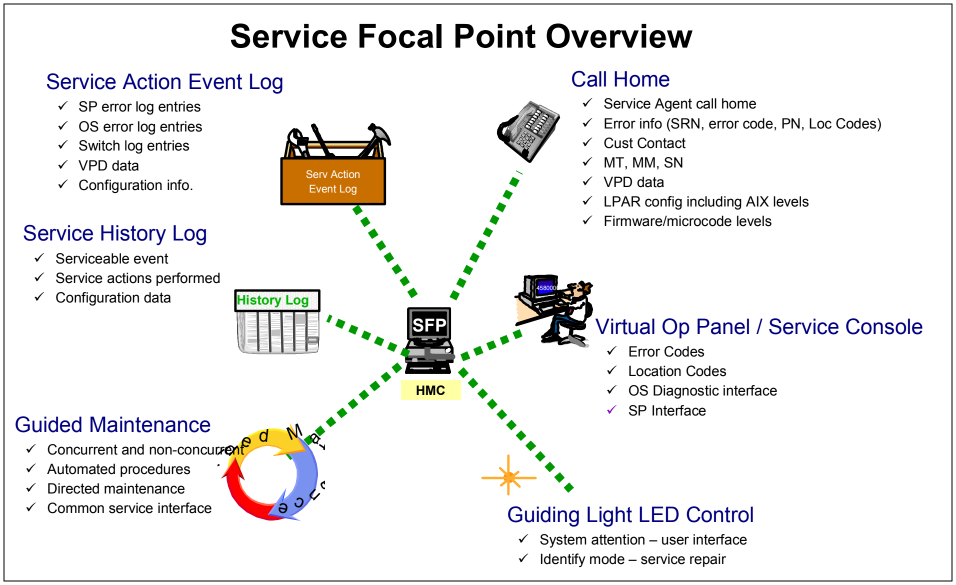
Service Agent can execute several tasks, including:

- Automatic problem analysis
- Problem-definable threshold levels for error reporting
- Automatic problem reporting
- Automatic client notification
- Visualize hardware error logs

By utilizing Service Agent, the p5-590 and p5-595 can reduce the amount of downtime experienced in the event of a system component failure by giving the service provider the ability to view the error report entry and, if needed, order any necessary replacement parts prior to arriving on site. The opportunity for human misinterpretation or miscommunication in problem determination is therefore mitigated.

6.6.6 Service Focal Point

Traditional service strategies become more complicated in a partitioned environment. Each partition runs on its own, unaware that other partitions exist on the same system. If one partition reports an error for a shared resource, such as a managed system power supply, other active partitions report the same error. To enable service representatives to avoid long lists of repetitive call-home information, the HMC provides the Service Focal Point application. Service Focal Point recognizes that these errors repeat, and filters them into one *serviceable event* for the service representative to review.



Service Focal Point Overview

Service Action Event Log
- ✓ SP error log entries
- ✓ OS error log entries
- ✓ Switch log entries
- ✓ VPD data
- ✓ Configuration info.

Call Home
- ✓ Service Agent call home
- ✓ Error info (SRN, error code, PN, Loc Codes)
- ✓ Cust Contact
- ✓ MT, MM, SN
- ✓ VPD data
- ✓ LPAR config including AIX levels
- ✓ Firmware/microcode levels

Service History Log
- ✓ Serviceable event
- ✓ Service actions performed
- ✓ Configuration data

Virtual Op Panel / Service Console
- ✓ Error Codes
- ✓ Location Codes
- ✓ OS Diagnostic interface
- ✓ SP Interface

Guided Maintenance
- ✓ Concurrent and non-concurrent
- ✓ Automated procedures
- ✓ Directed maintenance
- ✓ Common service interface

Guiding Light LED Control
- ✓ System attention – user interface
- ✓ Identify mode – service repair

Figure 6-8 Service focal point overview

As shown in Figure 6-8 on page 165, 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 allows a user to view the events and perform problem determination.

> **Note:** Service Focal Point only collects hardware errors, such as PERMANENT errors from AIX (marked as P) and NON BOOT errors from the service processor, and errors from the HMC itself.

The Service Focal Point (SFP) application is also the starting point of all service actions on HMC attached systems. The service representative begins the repair with the SFP application, selecting the Repair Serviceable Events view from the SFP Graphical User Interface (GUI). From here, the service representative selects a specific fault for repair from a list of open service events; initiating automated maintenance procedures specially designed for the IBM @server p5 and @server i5 systems.

While it is IBMs plan to provide automated service procedures for each component in the system, at first announce only those components that are concurrently maintainable are supported by the new automated processes. Additional components will be supported with future releases of firmware.

Automating various service procedural tasks, instead of relying on the service representative can help remove or significantly reduce the likelihood of service representative induced errors.

Many service tasks can be automated. For example, the HMC can guide the service representative to:

- Interpret error information.
- Prepare components for removal or initiate them after install.
- Set and reset system identify LEDs as part of the Guiding Light service approach.
- Automatically link to the next step in the service procedure based on input received from the current step.
- Update the service history log, indicating the service actions taken as part of the repair procedure. The history log helps to retain an accurate view of the service scenarios in case future actions are needed.

6.7 AIX RAS features

The RAS features described in the previous sections are all based on the p5-590 and p5-595 hardware. There are some additional RAS features that work in conjunction with the AIX operating system, and some that are entirely dependent on AIX. The following outlines some of the AIX RAS features.

- Unrecoverable error analysis

  AIX analyzes the hardware reported unrecoverable errors. If the process is a user process, it will be terminated. If the process is within the AIX kernel, then the operating system will be terminated. Again, terminating AIX in a partition will not impact any of the other partitions.

- Error log RAS

  Under very rare circumstances, such as powering off the system exactly while the errdaemon is writing into the error log, the error log may become corrupted. In AIX 5L Version 5.3 there are minor modifications made to the errdaemon to improve its robustness. The difference from the previous versions of AIX is that the errdaemon used to reset the log file if it was corrupted, instead of repairing it.

- System hang detection

  AIX 5L Version 5.1 and higher offers a feature called *system hang detection* that provides a mechanism to detect system hangs and initiates a pre-configured action. In some situations, it is difficult to distinguish a system that really hangs (it is not doing any meaningful work anymore) from a system that is so busy that none of the lower priority tasks, such as user shells, have a chance to run. The new system hang detection feature uses a shdaemon entry in the /etc/inittab file with an action field that specifies what should be done when certain conditions are met.

  For more information on how to configure this feature, refer to *Managing AIX Server Farms*, SG24-6606.

- AIX disk mirroring and LVM sparing

  Mirroring the operating system boot disks is a feature available on AIX since Version 4.2.1. It enables continued operation of the system, even in the case of failure on an operating system disk. Beginning with AIX 5L Version 5.1, it is possible to designate disks as hot spare disks in a volume group and to specify a policy to be used in the case of failing disks.

  For detailed information on rootvg mirroring on AIX, refer to *AIX Logical Volume Manager, From A to Z: Introduction and Concepts*, SG24-5432. For LVM hot spare disk support in a volume group, refer to *AIX 5L Differences Guide*, SG24-5765.

- TCP/IP RAS enhancements

  Beginning with AIX 5L Version 5.1, AIX provides several availability features in the TCP/IP network access.

  The *Dead Gateway Detection (DGD)* feature in AIX 5L Version 5.1 implements a mechanism for hosts to detect a dysfunctional gateway, adjust its routing table accordingly, and re-route network traffic to an alternate backup route, if available.

  With the *multi-path routing* feature in AIX 5L Version 5.1, routes no longer need to have a different destination, netmask, or group ID lists.

  These two features working together enable the system to route TCP/IP traffic through multiple routes and avoid the defective routes. This is as important as the server availability itself, because a network failure also causes an interrupt in server access.

  For Ethernet adapters, the *network interface backup* support allows multiple adapters to be grouped together and act as a single interface. In case of a failure in one interface, another interface will automatically take over the TCP/IP traffic and continue operation.

  For more information on how to configure this feature, refer to *Managing AIX Server Farms*, SG24-6606.

6.8 Linux RAS features

A key attribute of Linux on POWER is mission-critical RAS features. Drawing from IBMs autonomic computing efforts, IBM @server p5, pSeries and OpenPower continues to enhance the scope of its RAS capabilities.

The following s RAS features are supported when running Linux:

- Chipkill and ECC memory
- Disk mirroring (software)
- Journaled file system (several available under Linux)
- PCI Extended Error detection
- Redundant, hot-plug power and cooling (where available)
- Error reporting to Service Focal Point
- Error log analysis
- Boot-time processor and memory deallocation
- First Failure Data Capture
- service processor

Some of the POWER RAS features that are currently supported only with the Linux 2.6 kernel on POWER systems include:

- Hot-swapping of disk drives
- Dynamic Processor Deallocation
- Hot-plug PCI (future direction)
- PCI Extended Error recovery (future direction and device driver dependent)

To enable Linux to take advantage of the IBM @server p5, pSeries and OpenPower enhanced reliability support, a Service Aids Toolkit has been made available for download. This toolkit should greatly enhance Linux availability and serviceability when running on IBM @server p5, pSeries and OpenPower systems. The toolkit information and download is at the following URL:

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

7

Service processor

The service processor is the brains of the system, continuously monitoring and recording system functions and health.

Additional information on the service processor can be found in this publication in:

- Section 6.4.3, “Service processor and clocks” on page 147
- Chapter 8, “Hardware Management Console overview” on page 195

In this chapter you find the following information regarding the service processor:

7.1 Service processor functions

All IBM @server p5 590 and p5 595 servers are shipped with dual service processor cards. One is considered the primary and the other secondary.

They are connected together over the Ethernet ports in the bulk power controller (BPC).

It is IBMs intention to provide an automatic failover capability to the redundant service processor in 2005. IBM plans to make the failover support available through a no charge firmware upgrade.

> **Note:** All statements regarding IBM's plans, directions, and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

The service processor is responsible for performing several functions within the system. The purpose of this Chapter is to accomplish the following:

- Summarize how service processor binary image is packaged and executed on the platform
- Discuss how the service processor initiates the initial program load (IPL) for the platform
- Introduce how to use the Advanced System Management Interface (ASMI) for error reporting

7.1.1 Firmware binary image

Each service processor has two updatable non-volitile dedicated flash memory areas for storing the licensed internal code (LIC), also known as *firmware*. The sides are named the temporary *t* and a permanent *p* side. The t-side is where changes are generally applied and where the client generally runs from. The p-side should be viewed as a well known backup level that the client can revert to if necessary.

The CEC firmware binary image is a single image that includes code for the service processor, the POWER Hypervisor, and platform partition firmware. This CEC firmware binary image is stored on the system flash memory, as explained in Section 2.5, “System flash memory configuration” on page 34. The firmware image for the power code includes firmware for the bulk power controller (BPC), distributed converter assembly (DCA), fans, and clustering switch. The power code binary image is stored on the BPC service processor flash.

Since there are dual service processors per CEC, both service processors must be updated when firmware updates are applied and activated using the **Apply/Activate** command on the HMC.

The system power code image (stored on the BPC part of service processor) contains the power code for the frames that house the CEC cages, I/O cages, and clustering switches. The BPC service processor code load not only has the code load for the BPC service processor itself, but it also has the code for the DCA, bulk power regulator (BPR), fans, and other more granular field replaceable unit (FRUs) 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 may be installed in the frame.

The BPC part of the service processor has a two sided flash, and overall, it is the same hardware as the CEC service processor. When the concurrent firmware **Install/Activate** command from the HMC is initiated, the BPC service processor reboots as part of the concurrent activate process. The BPC initialization sequence central electronics complex after the reboot is unique. The BPC service processor must check the code levels of all the down level devices, including DCAs, BPRs, fans, cluster switches, and it must load those if they are different than what is in the active flash side of the BPC. Since the p5-590 and p5-595 have dual BPC service processors for each frame side, both will be updated as part of the firmware update process.

7.1.2 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.

7.1.3 Error Handling

Detection of certain CEC hardware failures will cause the primary service processor to terminate the IPL. After the IPL has been terminated the primary service processor will initiate a failover to the back-up service processor.

In the case that a system dump is performed, this will occur on the primary service processor detecting the CEC hardware failure. If the dump has not been extracted prior to the failover, then the CEC hardware data must be shadowed to the new primary. The CEC hardware data and main store memory would then be extracted from this new primary.

From the HMC a user can open an Advanced System Management (ASMI) browser session to view error log information and access the service processor. ASMI menus are accessible as long as there is the power applied to the frame. ASMI sessions can also be opened while partitions are running. It is possible to connect to the service processor using a personal computer as long as the system is connected to the service processor network. The IBM @server Hardware Information Center has a section on Accessing the Advanced System Management Interface.

7.2 Service processor cabling

There are five ports on the service processor card used for power (SPCN), light strip connectors, t bulk power controller (BPC) connectors and one unused connection (J05). Figure 7-1 on page 172 provides a front view of the service processor card connections. The service processor cable connections are summarized in Table 7-1 on page 173.

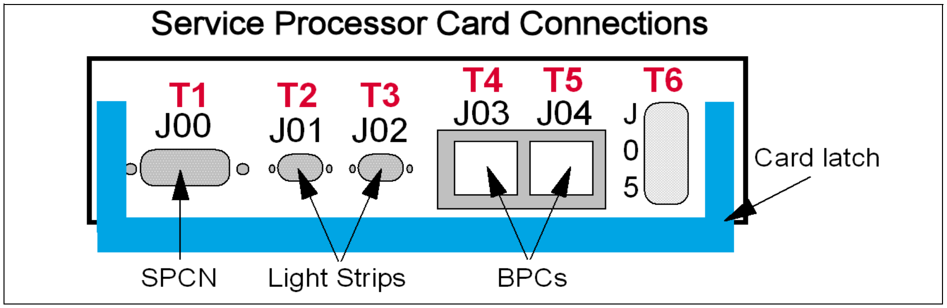


Figure 7-1 Service processor (front view)

Table 7-1 Table of service processor card location codes

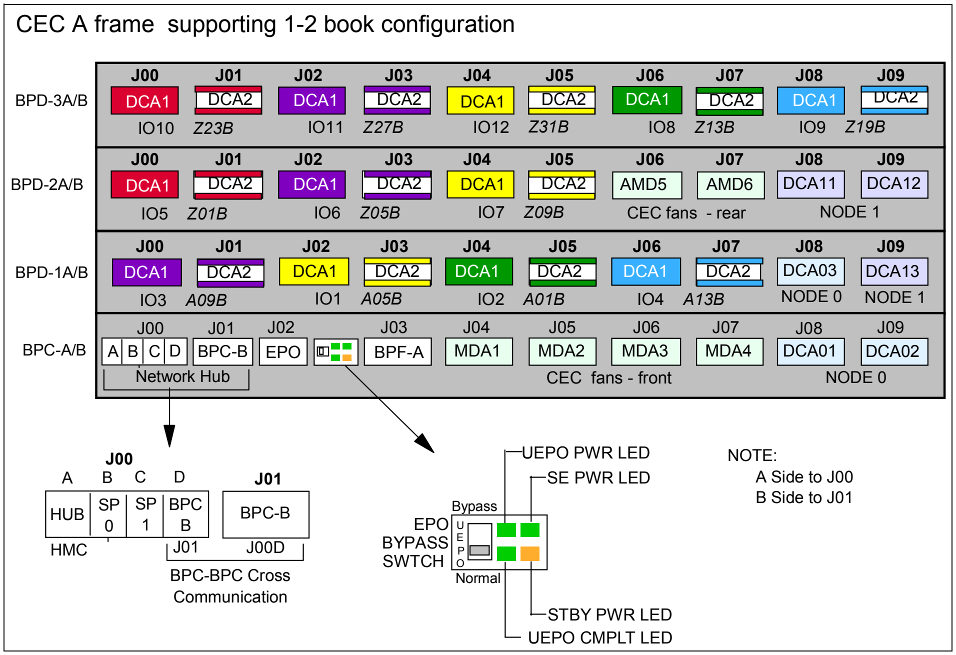
Jack ID	Location code	Service processor 0	Service processor 1	Function
J00	T1			System Power Control Network (SPCN) connection
J01	T2	J00 CEC Front Light Strip	J01 CEC Front Light Strip	Light Strip connection
J02	T3	J01 CEC Back Light Strip	J00 CEC Back Light Strip	
J03	T4	J00C BPA-A side	J00B BPA-A side	Ethernet port 0 to Bulk Power Controller (BPC)
J04	T5	J00C BPA-B side	J00B BPA-B side	Ethernet port 1 to Bulk Power Controller (BPC)
J05	T6	Unused		

The p5-590 and p5-595 are designed with dual bulk power controllers (BPC). Each bulk power controller BPC) has two sides, commonly referred to as A and B sides. Additionally, there are four 10/100 Ethernet hub ports on the BPC that connect to various system components. This connectivity scheme is presented in Table 7-2 on page 173. Figure 7-2 on page 174 illustrates the available connectors on each bulk power controller (BPC).

Figure 7-3 on page 174 provides a full illustration of the service processor card, clock oscillator, and bulk power controller (BPC) integration.

Table 7-2 Summary of BPC Ethernet hub port connectors

BPC Ethernet hub port	Connected component
BPC Port A	Connects to the Hardware Management Console (HMC)
BPC Port B	Connects to service processor 0
BPC Port C	Connects to service processor 1
BPC Port D	Connects to the partner BPC



CEC A frame supporting 1-2 book configuration

	J00	J01	J02	J03	J04	J05	J06	J07	J08	J09
BPD-3A/B	DCA1	DCA2	DCA1	DCA2	DCA1	DCA2	DCA1	DCA2	DCA1	DCA2
	IO10	Z23B	IO11	Z27B	IO12	Z31B	IO8	Z13B	IO9	Z19B
BPD-2A/B	DCA1	DCA2	DCA1	DCA2	DCA1	DCA2	AMD5	AMD6	DCA11	DCA12
	IO5	Z01B	IO6	Z05B	IO7	Z09B	CEC fans - rear		NODE 1	
BPD-1A/B	DCA1	DCA2	DCA1	DCA2	DCA1	DCA2	DCA1	DCA2	DCA03	DCA13
	IO3	A09B	IO1	A05B	IO2	A01B	IO4	A13B	NODE 0	NODE 1
BPC-A/B	A B C D	BPC-B	EPO 🟩🟩🟩🟧	BPF-A	MDA1	MDA2	MDA3	MDA4	DCA01	DCA02
	Network Hub				CEC fans - front				NODE 0	

Figure 7-2 Bulk power controller connections

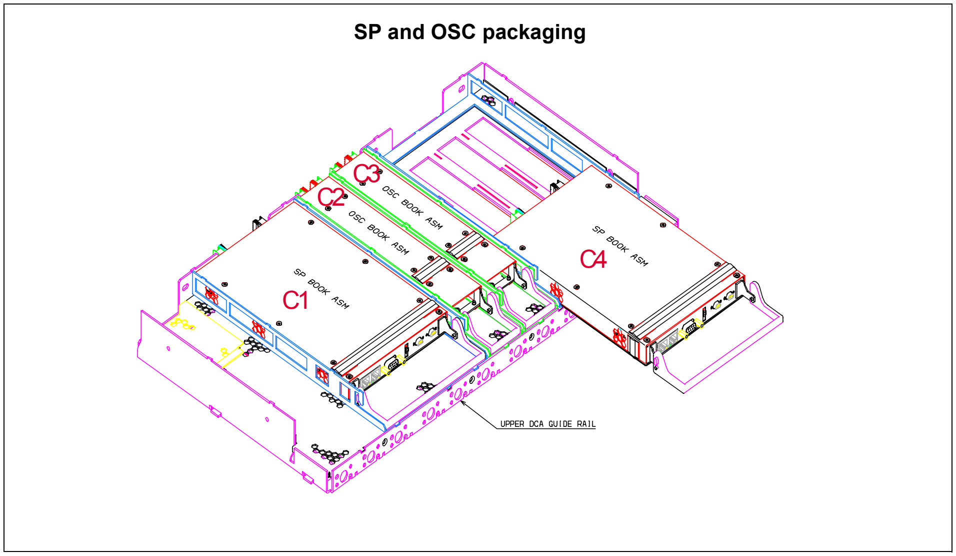


Figure 7-3 Oscillator and service processor

7.3 Advanced System Management Interface (ASMI)

The Web interface to the Advanced System Management Interface (ASMI) 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 (IPL) 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.

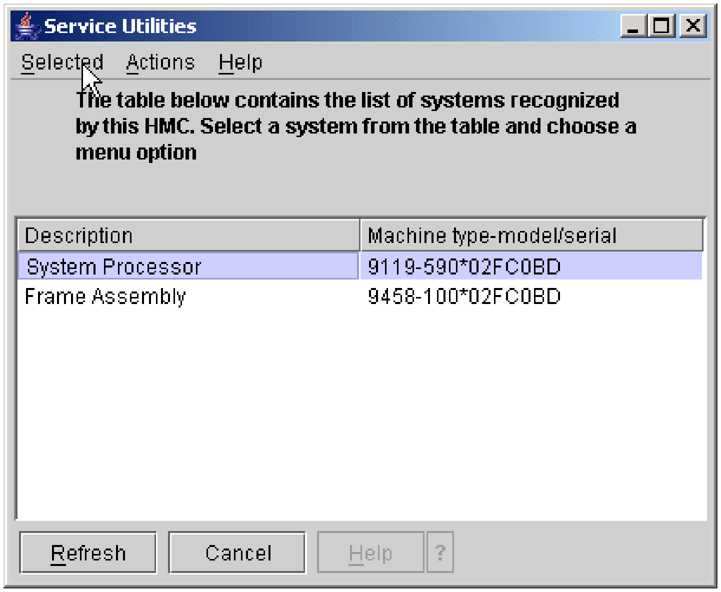
The purpose of this section is to provide screen shots as reference for configuring Advanced System Management Interface (ASMI) menus. This section has two main components:

- How to access the ASMI menus using the HMC
- How to access the ASMI menus using a Web browser
- Summarize ASMI User accounts
- Overview of various ASMI functions.

7.3.1 Accessing ASMI using HMC Service Focal Point utility

The HMC Service Focal Point application is the entry point into gaining ASMI access. On the HMC Service Focal Point application, select the **Service Utilities** option. Figure 7-4 on page 176 provides a screen shot of the Service Utilities menu.

From the Service Utilites menu, you must select the appropriate machine type-mode/serial number. This information is highlighted in blue in Figure 7-4 on page 176.



The table below contains the list of systems recognized by this HMC. Select a system from the table and choose a menu option

Description	Machine type-model/serial
System Processor	9119-590*02FC0BD
Frame Assembly	9458-100*02FC0BD

Figure 7-4 Select service processor

Then, click on the `Selected` pull-down menu and click on Launch ASM Menu, as illustrated in Figure 7-5 on page 176.

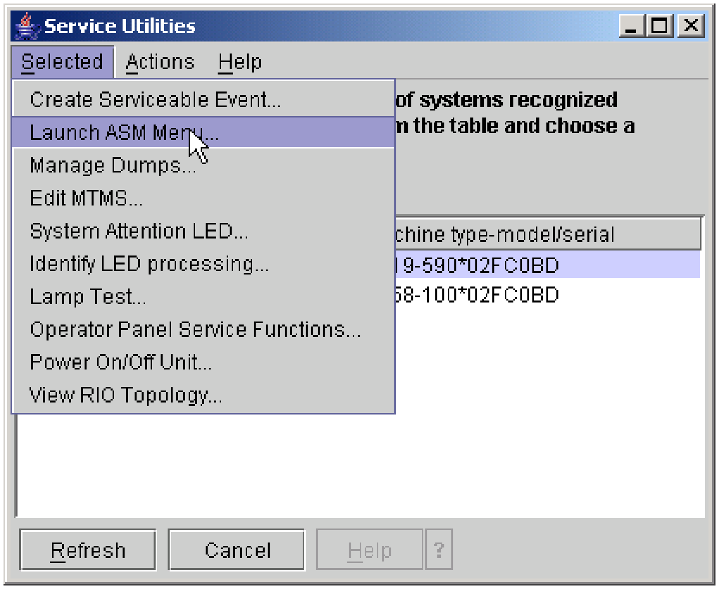
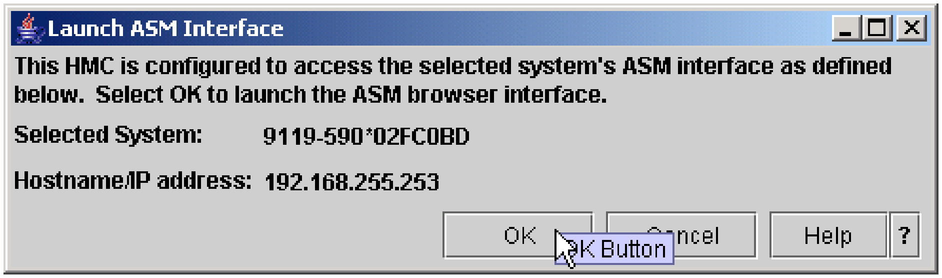


Figure 7-5 Select ASMI

As part of the Launch ASM Menu dialog, the serial number and hostname/IP address of the service processor are presented and must be verified, as show in Figure 7-6.



Launch ASM Interface

This HMC is configured to access the selected system's ASM interface as defined below. Select OK to launch the ASM browser interface.

Selected System: 9119-590*02FC0BD

Hostname/IP address: 192.168.255.253

Figure 7-6 OK to launch

7.3.2 Accessing ASMI using a Web browser

The following instructions apply to systems that are not connected to a Hardware Management Console (HMC), or if you have the 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.

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.

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.

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.

Determine the IP address of the Ethernet port that your PC or mobile computer is connected to, and enter the IP address in the address field of the Web browser.

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

```
https://192.168.2.147
```

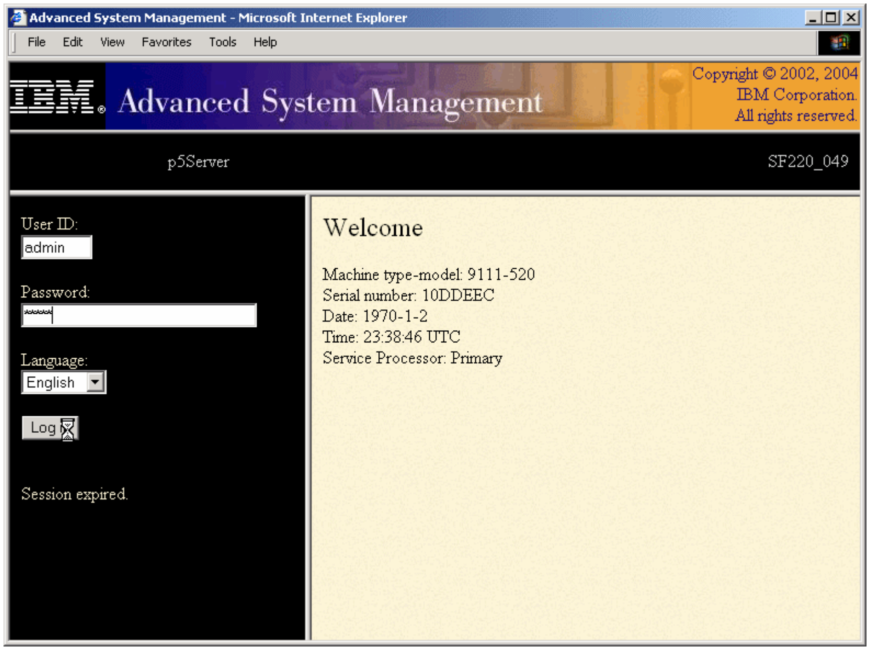
When the login display appears, enter one of the user accounts and passwords provided in Table 7-3.

7.3.3 ASMI login screen

Once the ASMI menus are successfully launched, the user will be presented with a login screen. The user login screen is divided into two sides: left and right. The left hand side contains menu options for conducting certain system operations. The menu options on the left hand side are static and constantly presented to the user. The right hand side contains dynamic data that changes based upon the menu option the user has chosen.

The right hand side is populated with a Welcome screen, system machine type-model, system serial number, current date and time, and identifies which service processor the system is connected to (primary or secondary). A list of current users and their associated IP addresses are also presented in the welcome screen. This login screen will also reveal the licensed internal code (LIC) version in the upper right corner, as show in Figure 7-7 on page 178.

.


Advanced System Management - Microsoft Internet Explorer
File Edit View Favorites Tools Help
IBM Advanced System Management
Copyright © 2002, 2004 IBM Corporation. All rights reserved.
p5Server
SF220_049
Log In

User ID: admin

Password: *****

Language: English

Session expired.

Welcome

Machine type-model: 9111-520
Serial number: 10DDEEC
Date: 1970-1-2
Time: 23:38:46 UTC
Service Processor: Primary

Figure 7-7 ASMI login

7.3.4 ASMI user accounts

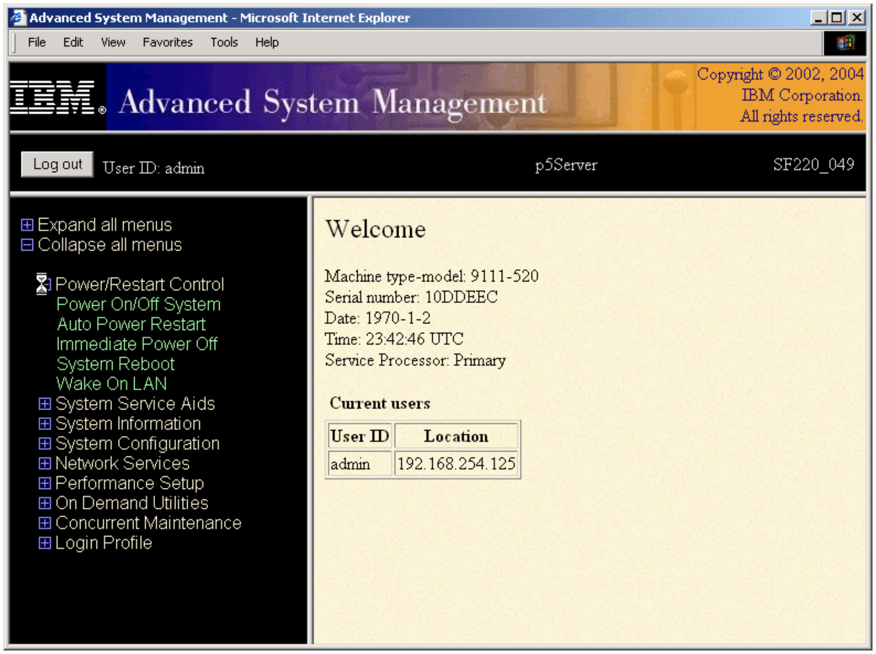
In Table 7-3, you find the default user login and passwords. Usually, you will use the ‘admin login, however, be aware that the client has the ability to change any password.

Table 7-3 ASMI user accounts

User ID	Password
admin	admin
general	general

Once a user has successfully logged in to the ASMI menus, there are various tasks which can be executed, which are highlighted in the next section.

There are various levels of ASMI user access depending upon the user login, and the menu options on the left hand side of the ASMI screen are different for general and admin users. Admin users have a more comprehensive set of functions available, as shown in Table 7-4 on page 180. Each menu option has several sub-menu option available. Figure 7-8 on page 179 presents a user who has successfully logged into the system as admin.



Welcome

Machine type-model: 9111-520
Serial number: 10DDEEC
Date: 1970-1-2
Time: 23:42:46 UTC
Service Processor: Primary

Current users

User ID	Location
admin	192.168.254.125

Figure 7-8 ASMI menu: Welcome (as admin)

7.3.5 ASMI menu functions

The purpose of this section is to provide a snapshot of the ASMI capabilities for admin level users. Every menu option is not covered within this publication.

Table 7-4 ASMI user-level access (menu options)

Admin user level access	General user level access
► Power/Restart Control ► System Service Aids ► System Information ► System Configuration ► Network Services ► Performance Setup ► On Demand Utilities ► Concurrent Maintenance ► Login Profile	► System Service Aids ► System Configuration ► Login Profile

7.3.6 Power On/Off tasks

As an admin user, the following power-on parameters can be defined using the Power On/Off System ASMI menu option:

- Select boot mode: Fast/Slow
- Select firmware side: Temporary/Permanent
- Select service mode: Normal/Manual
- Select boot state: Standby/Running

7.3.7 System Service Aids tasks

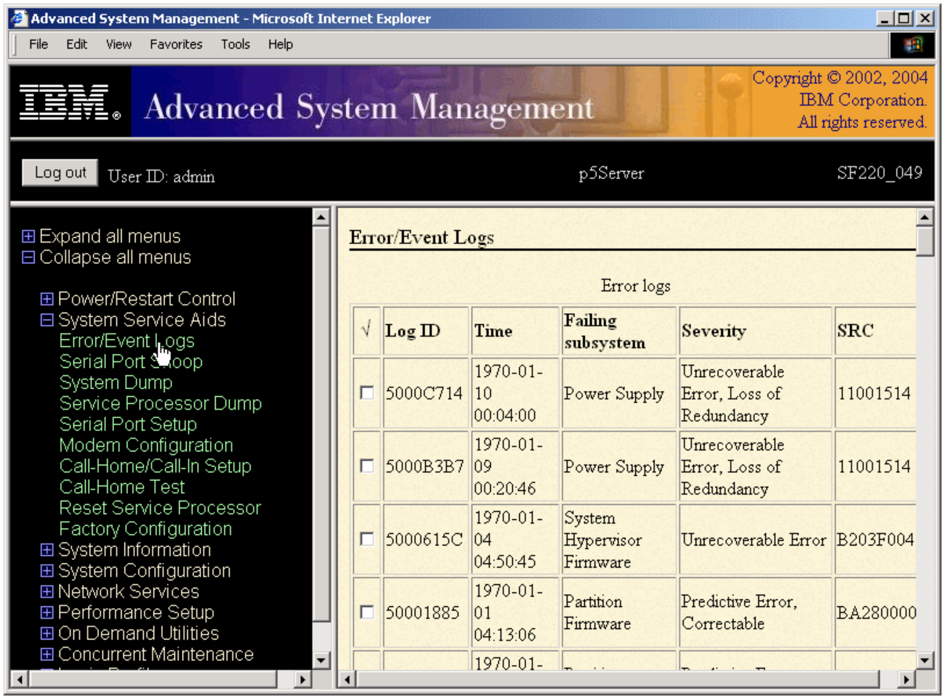
As an admin user, the following power-on parameters can be defined using the System Service Aids ASMI menu option:

- Service processor error log
- Reset service processor

  Be aware that the ASMI Reset service processor menu function is comparable to the reset with the pinhole in the operate panel. This is used to recover from an unresponsive service processor.
- Factory Configuration

Service processor error log

Figure 7-9 on page 181 provides an example of the service processor error log. You can click on each individual event to obtain a more detailed view of the error.



Error/Event Logs

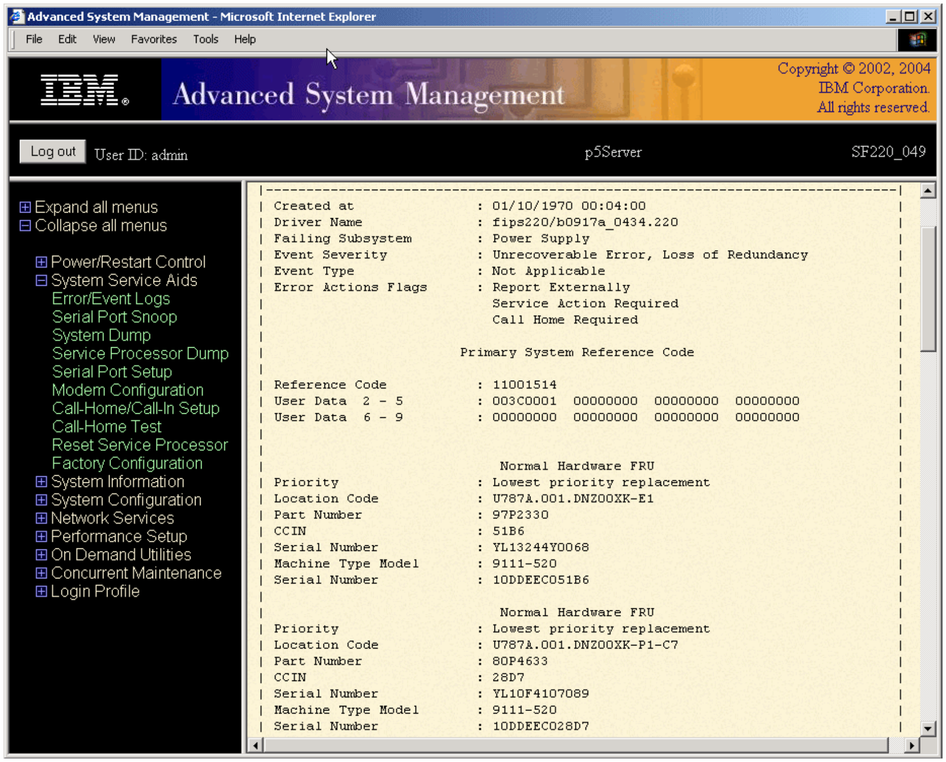
Error logs

√	Log ID	Time	Failing subsystem	Severity	SRC
□	5000C714	1970-01-10 00:04:00	Power Supply	Unrecoverable Error, Loss of Redundancy	11001514
□	5000B3B7	1970-01-09 00:20:46	Power Supply	Unrecoverable Error, Loss of Redundancy	11001514
□	5000615C	1970-01-04 04:50:45	System Hypervisor Firmware	Unrecoverable Error	B203F004
□	50001885	1970-01-01 04:13:06	Partition Firmware	Predictive Error, Correctable	BA280000
		1970-01-			

Figure 7-9 ASMI menu: Error /Event Logs

Detailed service processor error log

Figure 7-10 on page 182 provides a detailed view of the error log. Note that it resembles the operating system error log.

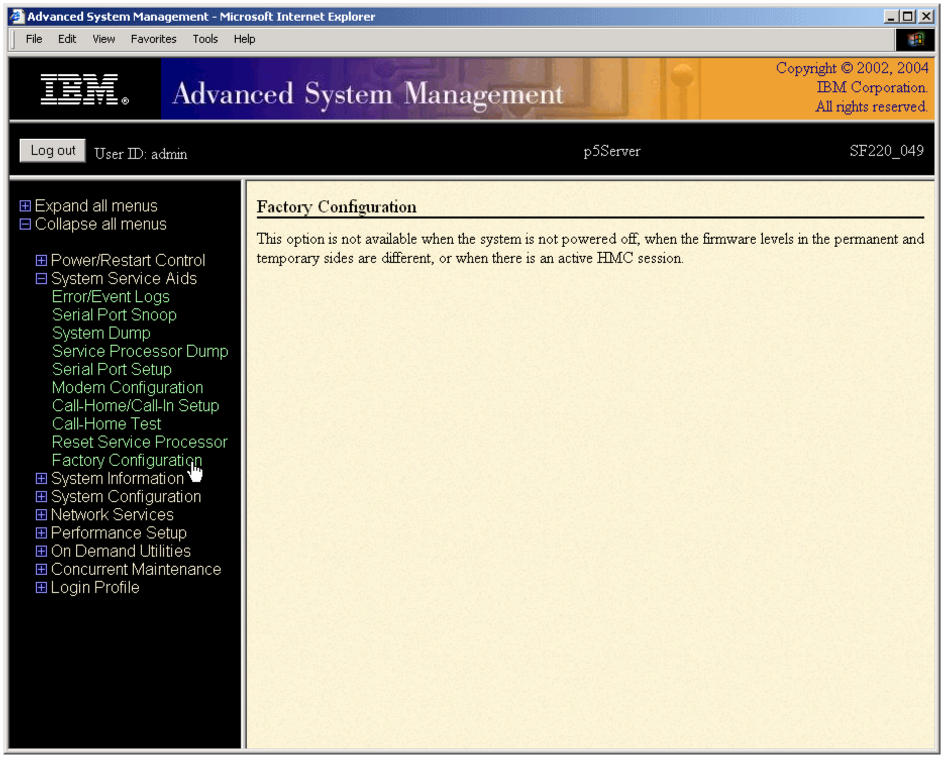


```
|---------------------------------------------------------------------------|
| Created at                : 01/10/1970 00:04:00                          |
| Driver Name               : fips220/b0917a_0434.220                      |
| Failing Subsystem         : Power Supply                                 |
| Event Severity            : Unrecoverable Error, Loss of Redundancy      |
| Event Type                : Not Applicable                               |
| Error Actions Flags       : Report Externally                            |
|                             Service Action Required                      |
|                             Call Home Required                           |
|                                                                           |
|                          Primary System Reference Code                    |
|                                                                           |
| Reference Code            : 11001514                                     |
| User Data  2 - 5          : 003C0001  00000000  00000000  00000000       |
| User Data  6 - 9          : 00000000  00000000  00000000  00000000       |
|                                                                           |
|                                                                           |
|                               Normal Hardware FRU                         |
| Priority                  : Lowest priority replacement                  |
| Location Code             : U787A.001.DNZ00XK-E1                         |
| Part Number               : 97P2330                                      |
| CCIN                      : 51B6                                         |
| Serial Number             : YL13244Y0068                                 |
| Machine Type Model        : 9111-520                                     |
| Serial Number             : 10DDEEC051B6                                 |
|                                                                           |
|                               Normal Hardware FRU                         |
| Priority                  : Lowest priority replacement                  |
| Location Code             : U787A.001.DNZ00XK-P1-C7                      |
| Part Number               : 80P4633                                      |
| CCIN                      : 28D7                                         |
| Serial Number             : YL10F4107089                                 |
| Machine Type Model        : 9111-520                                     |
| Serial Number             : 10DDEEC028D7                                 |
```

Figure 7-10 ASMI menu: Detailed Error Log

Factory Configuration

The Factory Configuration menu (Figure 7-11) is used to reset the service processor back to its default factory configuration. Resetting the service processor using this menu option deletes all configuration data such as network configuration information.



Factory Configuration

This option is not available when the system is not powered off, when the firmware levels in the permanent and temporary sides are different, or when there is an active HMC session.

Figure 7-11 ASMI menu: Factory Configuration

7.3.8 System Configuration ASMI menu

The ASMI System Configuration menu contains options to view or modify system-level information including the system name, processing unit identifier, firmware update policy, and time of day. These menu options are visible on the right hand side of Figure 7-11 on page 183. This section will provide screen shots for the firmware update policy menu. Other useful ASMI menu options are summarized below, and the content of the screens is left as an exercise for you.

- Configure I/O Enclosures

  With this dialog, you can change the settings of a I/O drawer. The ASMI menu option will reveal the serial number of the I/O drawer.

- Program Indicator History

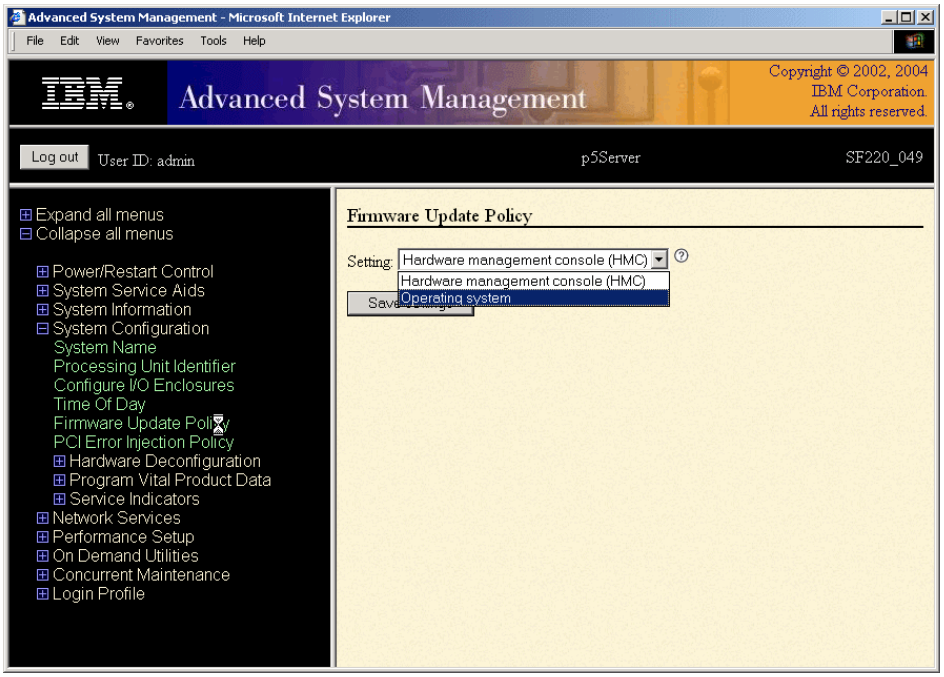
  With the program indicator from the last boot, you can check the boot process.

- Program Vital Product Data

  With this menu, you can access all of the managed system's vital product data (VPD).

Firmware Update Policy

A typical customization is to define the firmware licensed internal code (LIC) update policy. This setting specifies whether firmware updates will be done by the HMC or by the operating system. The ASMI menus provide an option to establish this policy, as shown in Figure 7-12.



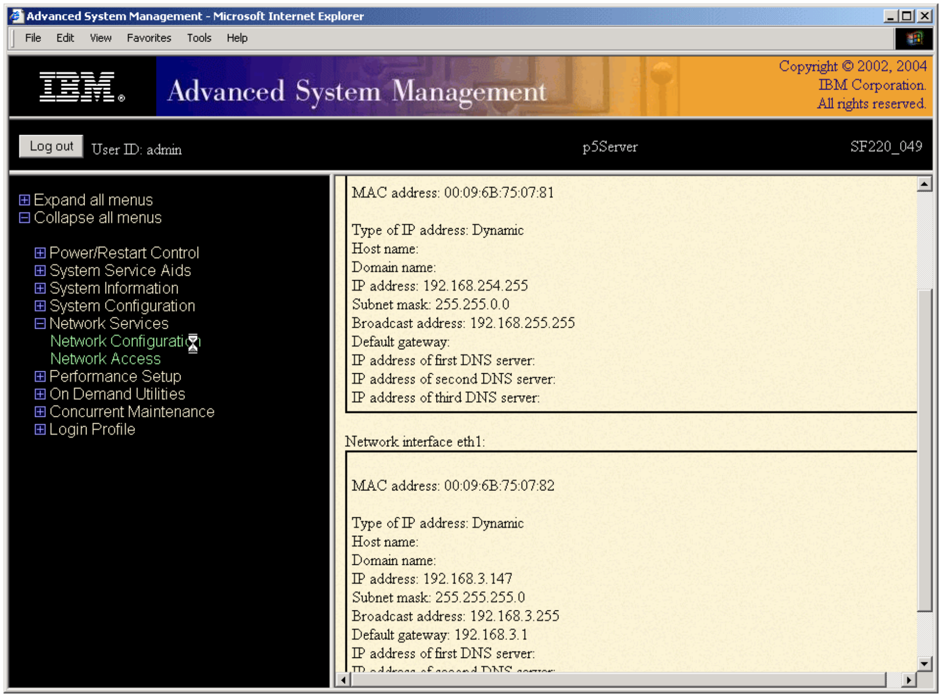
Setting: Hardware management console (HMC)

Figure 7-12 ASMI Menu: Firmware Update Policy

7.3.9 Network Services ASMI menu

The Network Services menu contains options to view and modify the network configuration and network access.

Figure 7-13 shows the Network Configuration menu. This menu allows a user to change the network configuration. However, changes to the system network configuration must be done when the system is powered off.



MAC address: 00:09:6B:75:07:81

Type of IP address: Dynamic
Host name:
Domain name:
IP address: 192.168.254.255
Subnet mask: 255.255.0.0
Broadcast address: 192.168.255.255
Default gateway:
IP address of first DNS server:
IP address of second DNS server:
IP address of third DNS server:

Network interface eth1:

MAC address: 00:09:6B:75:07:82

Type of IP address: Dynamic
Host name:
Domain name:
IP address: 192.168.3.147
Subnet mask: 255.255.255.0
Broadcast address: 192.168.3.255
Default gateway: 192.168.3.1
IP address of first DNS server:
IP address of second DNS server:

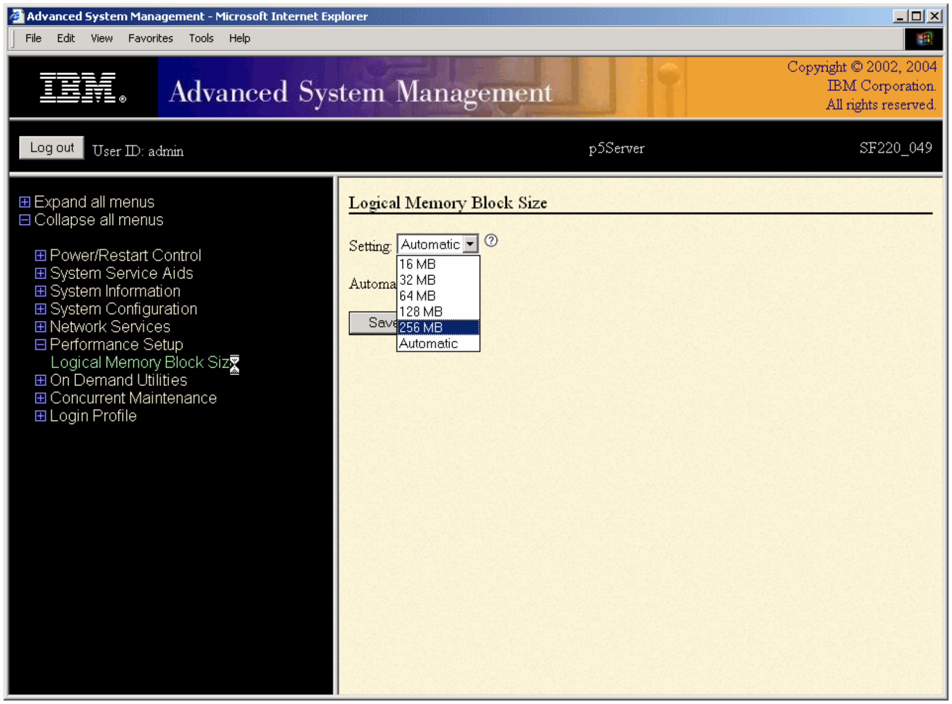
Figure 7-13 ASMI menu: Network Configuration

7.3.10 Performance Setup ASMI menu

The ASMI Performance Setup menu contains options to view and define the logical memory block size.

ASMI allows the user to select the desired logical memory block size, as shown in Figure 7-14 on page 186. The *Automatic* setting instructs the service processor to compute a value based on the amount of available memory. Memory block updates do not take effect until next system reboot. The selected size is the minimum size which the client can change after the next reboot in

DLPAR, memory resources. Logical memory blocks are also discussed in Section 5.2.4, “Memory configuration rules” on page 121.



Copyright © 2002, 2004
IBM Corporation.
All rights reserved.

User ID: admin p5Server SF220_049

Logical Memory Block Size

Setting: Automatic

16 MB
32 MB
64 MB
128 MB
256 MB
Automatic

Automa

Figure 7-14 ASMI menu: Logical Memory Block Size

7.4 Firmware updates

IBM will periodically release firmware updates for the p5-590 and p5-595. 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.

You must manage several types of fixes to properly maintain your hardware. Figure 7-15 shows the different types of hardware and software that might require fixes.

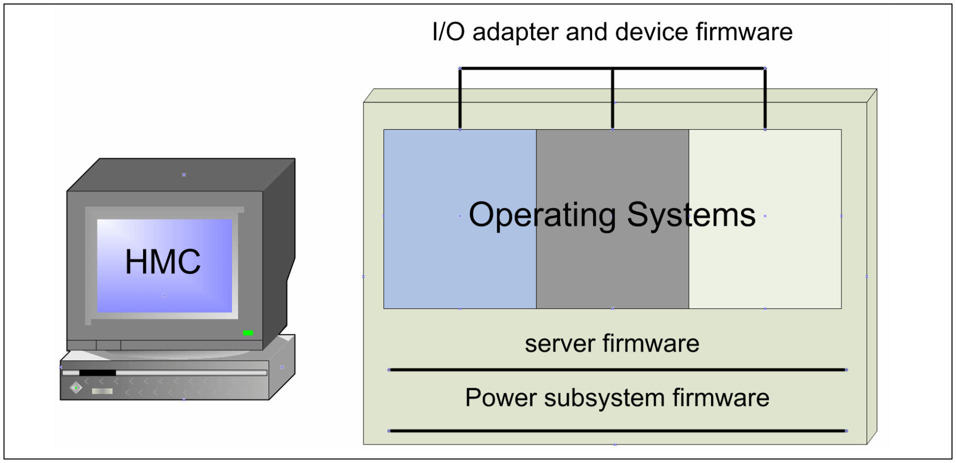
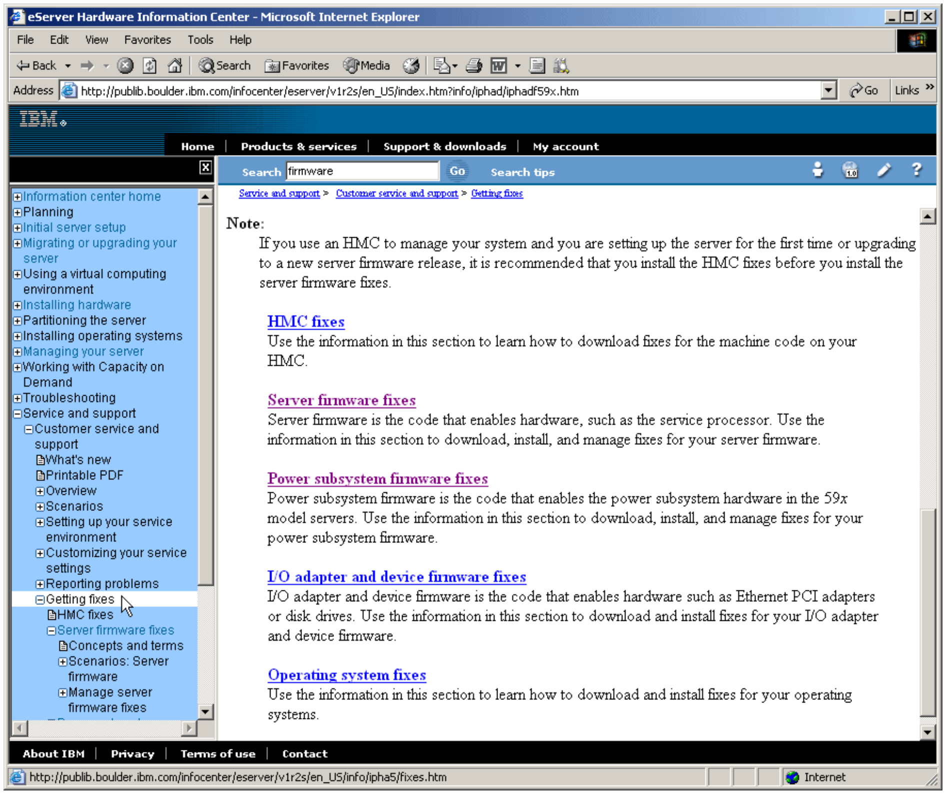


Figure 7-15 Potential system components that require fixes

Detailed firmware updating information can be found in the IBM @server Hardware Information Center under “Service and support > Customer service and support > Getting fixes” as shown on in Figure 7-16. Appendix D, “System documentation” on page 265 explains how to access the IBM @server Hardware Information Center.


eServer Hardware Information Center - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://publib.boulder.ibm.com/infocenter/eserver/v1r2s/en_US/index.htm?info/iphad/iphadf59x.htm
Go Links
IBM
Home | Products & services | Support & downloads | My account
Search firmware Go Search tips
Information center home
Planning
Initial server setup
Migrating or upgrading your server
Using a virtual computing environment
Installing hardware
Partitioning the server
Installing operating systems
Managing your server
Working with Capacity on Demand
Troubleshooting
Service and support
Customer service and support
What's new
Printable PDF
Overview
Scenarios
Setting up your service environment
Customizing your service settings
Reporting problems
Getting fixes
HMC fixes
Server firmware fixes
Concepts and terms
Scenarios: Server firmware
Manage server firmware fixes
Service and support > Customer service and support > Getting fixes
About IBM | Privacy | Terms of use | Contact
http://publib.boulder.ibm.com/infocenter/eserver/v1r2s/en_US/info/ipha5/fixes.htm
Internet

Note:

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, it is recommended that you install the HMC fixes before you install the server firmware fixes.

HMC fixes

Use the information in this section to learn how to download fixes for the machine code on your HMC.

Server firmware fixes

Server firmware is the code that enables hardware, such as the service processor. Use the information in this section to download, install, and manage fixes for your server firmware.

Power subsystem firmware fixes

Power subsystem firmware is the code that enables the power subsystem hardware in the 59x model servers. Use the information in this section to download, install, and manage fixes for your power subsystem firmware.

I/O adapter and device firmware fixes

I/O adapter and device firmware is the code that enables hardware such as Ethernet PCI adapters or disk drives. Use the information in this section to download and install fixes for your I/O adapter and device firmware.

Operating system fixes

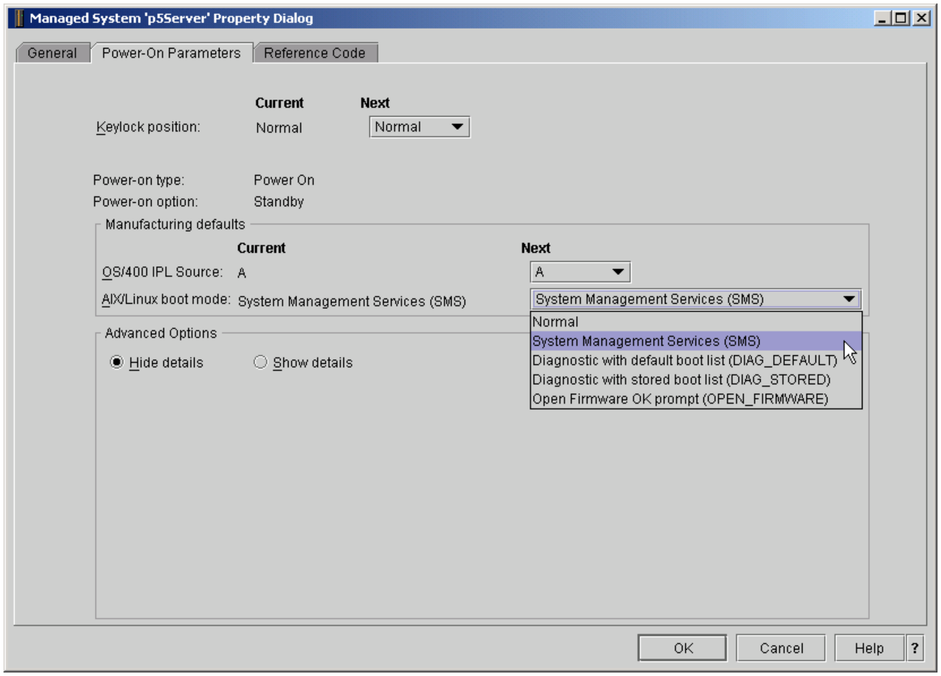
Use the information in this section to learn how to download and install fixes for your operating systems.

Figure 7-16 Getting fixes from the IBM @server Hardware Information Center

7.5 System Management Services

On the p5-590 and p5-595, the System Management Services (SMS) menus are used to specify the partition boot device. In order to change the boot device options, the user must boot the partition to the System Management Services (SMS) menu.

In order to boot to the SMS menu, the user must modify the power-on properties for the partition profile using the HMC. The user should select the pull-down menu for the partition boot mode and select the System Management Services (SMS) option. In Figure 7-17 on page 189, the partition is an AIX/Linux partition, and the user has selected the SMS boot mode option.


Managed System 'p5Server' Property Dialog

General
Power-On Parameters
Reference Code

	Current	Next
Keylock position:	Normal	Normal

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

Manufacturing defaults

	Current	Next
OS/400 IPL Source:	A	A
AIX/Linux boot mode:	System Management Services (SMS)	System Management Services (SMS)

Normal
System Management Services (SMS)
Diagnostic with default boot list (DIAG_DEFAULT)
Diagnostic with stored boot list (DIAG_STORED)
Open Firmware OK prompt (OPEN_FIRMWARE)

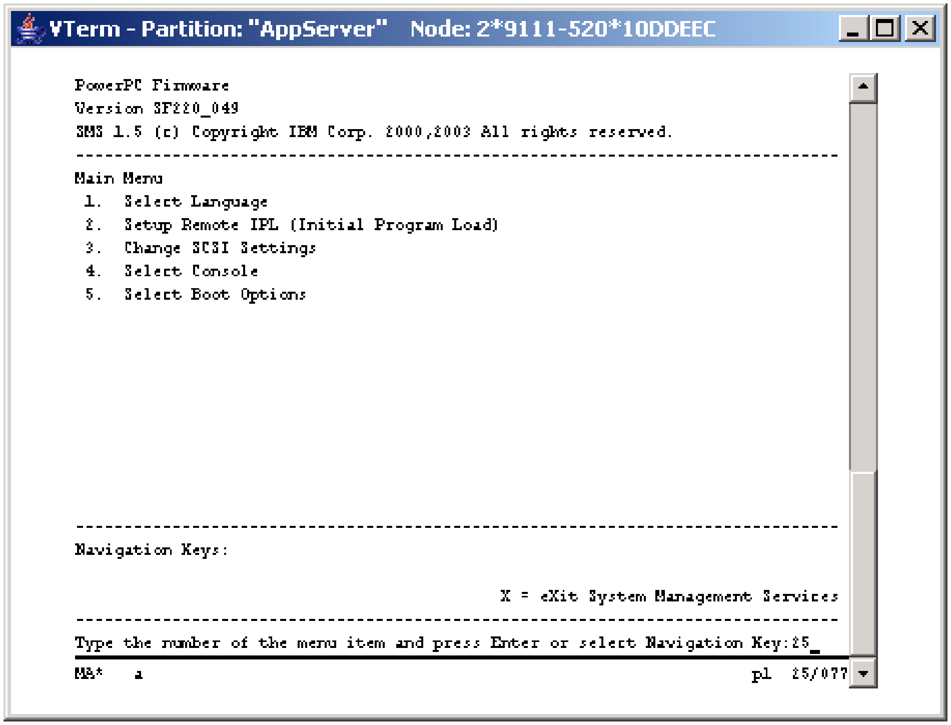
Advanced Options

● Hide details ○ Show details

Figure 7-17 Partition profile power-on properties

After the power-on properties in the partition profile have been updated, the boot sequence will pause once the SMS menu is reached. The user will be presented with the SMS main menu, as seen in Figure 7-18 on page 190. The SMS main menu has options for language selection, remote IPL setup, change SCSI settings, select console options, and select boot options.

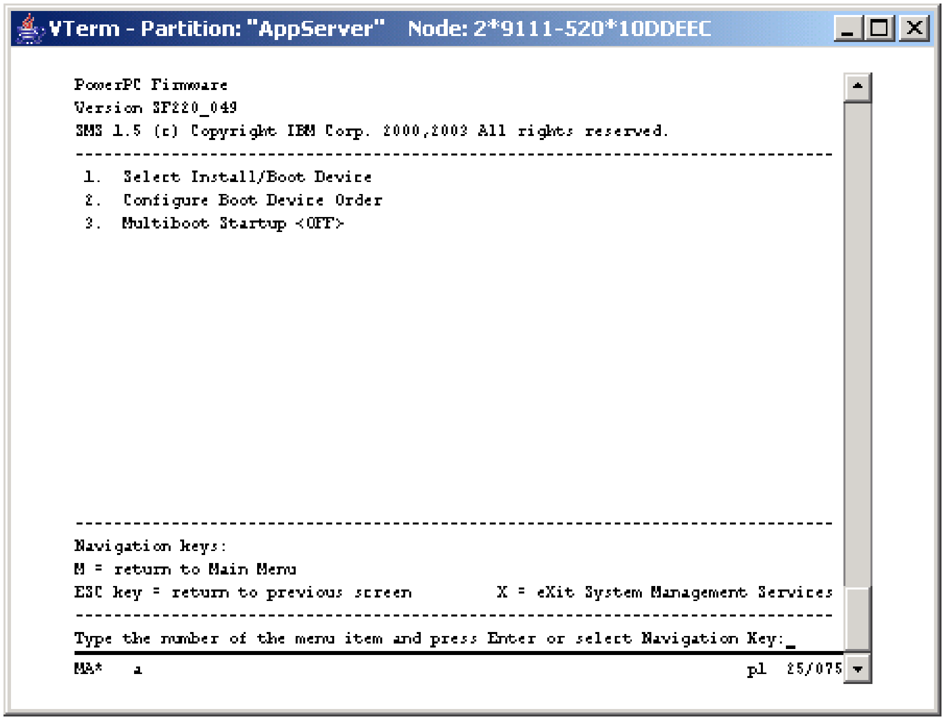
The screen captures are shown from a 9111-520 (p5-520) server, but the layout and function is identical on the p5-590 and p5-595.



```
PowerPC Firmware
Version SF220_049
SMS 1.5 (c) Copyright IBM Corp. 2000,2003 All rights reserved.
-------------------------------------------------------------------------------
Main Menu
 1.   Select Language
 2.   Setup Remote IPL (Initial Program Load)
 3.   Change SCSI Settings
 4.   Select Console
 5.   Select Boot Options








-------------------------------------------------------------------------------
Navigation Keys:

                                             X = eXit System Management Services
-------------------------------------------------------------------------------
Type the number of the menu item and press Enter or select Navigation Key:25_
MA*   a                                                              pl  25/077
```

Figure 7-18 System Management Services (SMS) main menu

Figure 7-19 on page 191 presents the user with three options to define the boot sequence. The user has the option to select a specific boot device, configure the boot device order, and toggle the multiboot startup state. Note that if multiboot is switched ON, then the LPAR will always stop at the SMS menu (to allow selection of the boot device), with the Operator Panel (on the HMC) showing either CA00E1DC (to select console) or CA00E1AE (to select boot option). Note that if the user intends to install the boot image from CD or boot from a diagnostic CD, the user should select the Select Install/Boot Device menu option.

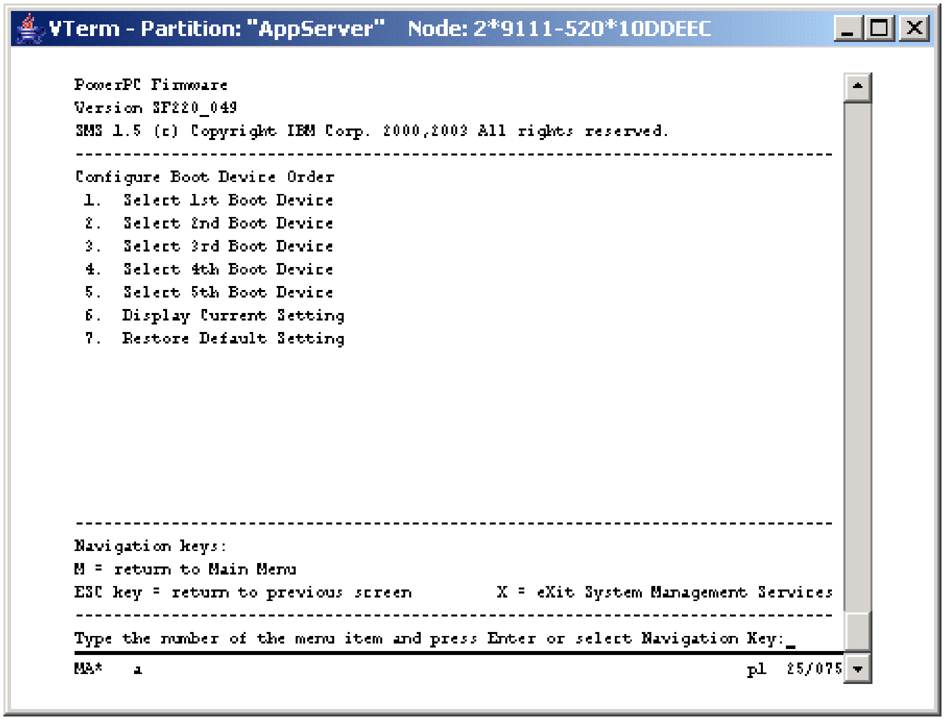


```
PowerPC Firmware
Version SF220_049
SMS 1.5 (c) Copyright IBM Corp. 2000,2003 All rights reserved.
-------------------------------------------------------------------------------
 1.  Select Install/Boot Device
 2.  Configure Boot Device Order
 3.  Multiboot Startup <OFF>









-------------------------------------------------------------------------------
Navigation keys:
M = return to Main Menu
ESC key = return to previous screen         X = eXit System Management Services
-------------------------------------------------------------------------------
Type the number of the menu item and press Enter or select Navigation Key:_
MA*   a                                                              p1  25/075
```

Figure 7-19 Select Boot Options menu options

Figure 7-20 on page 192 shows the Configure Boot Device Order menu options. The boot device order will be stored for the subsequent system boots.

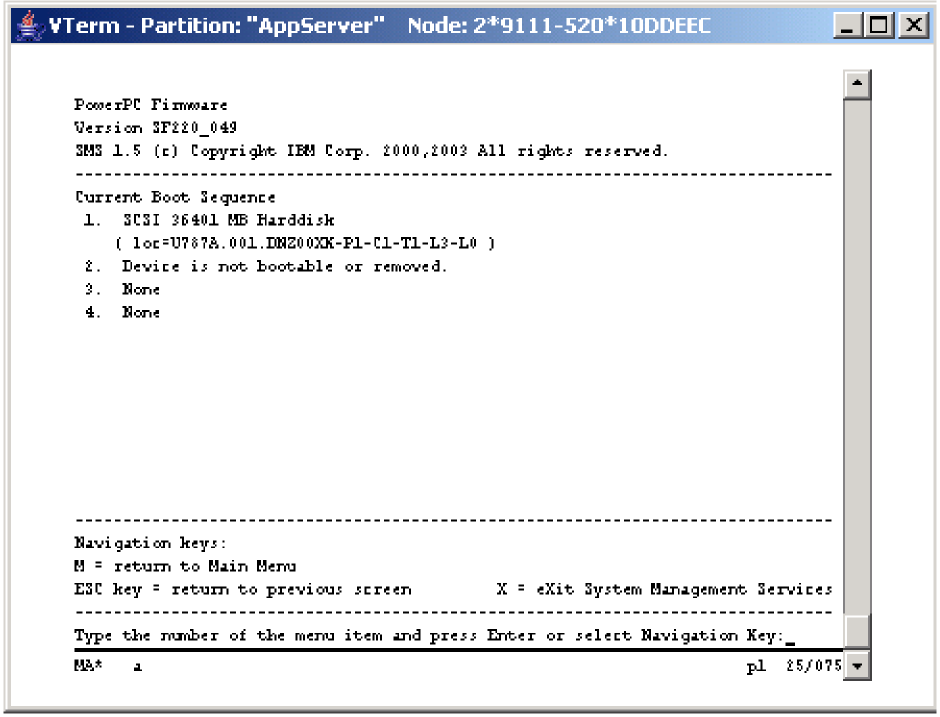
> **Note:** The user must check the default boot list before changing the boot list order.



```
PowerPC Firmware
Version SF220_049
SMS 1.5 (c) Copyright IBM Corp. 2000,2003 All rights reserved.
-------------------------------------------------------------------------------
Configure Boot Device Order
 1.   Select 1st Boot Device
 2.   Select 2nd Boot Device
 3.   Select 3rd Boot Device
 4.   Select 4th Boot Device
 5.   Select 5th Boot Device
 6.   Display Current Setting
 7.   Restore Default Setting









-------------------------------------------------------------------------------
Navigation keys:
M = return to Main Menu
ESC key = return to previous screen         X = eXit System Management Services
-------------------------------------------------------------------------------
Type the number of the menu item and press Enter or select Navigation Key:_
MA*    a                                                                p1  25/075
```

Figure 7-20 Configure Boot Device Order menu

Figure 7-21 on page 193 provides an example of a default boot list. If the default boot list is empty, the user must return to the Configure Boot Device Order menu and select the Restore Default Settings option.



```
PowerPC Firmware
Version SF220_049
SMS 1.5 (c) Copyright IBM Corp. 2000,2003 All rights reserved.
-------------------------------------------------------------------------------
Current Boot Sequence
 1.   SCSI 36401 MB Harddisk
    ( loc=U787A.001.DNZ00XK-P1-C1-T1-L3-L0 )
 2.   Device is not bootable or removed.
 3.   None
 4.   None










-------------------------------------------------------------------------------
Navigation keys:
M = return to Main Menu
ESC key = return to previous screen         X = eXit System Management Services
-------------------------------------------------------------------------------
Type the number of the menu item and press Enter or select Navigation Key:_
MA*   a                                                              pl  25/075
```

Figure 7-21 Current boot sequence menu (default boot list)

8

Hardware Management Console overview

A Hardware Management Console (HMC) is a desktop or rack mounted computer, very similar to the kind that most of us use every day. What makes an HMC different from other personal computers is that the HMC runs a unique combination of software applications, including proprietary software that provides the HMC the ability to manage IBM eserver systems.

8.1 Introduction

All p5-590 and p5-595 configurations require the use of a Hardware Management Console (HMC) to manage hardware and low-level software functions.

There are some significant differences regarding the HMC connection to the managed server with the p5-590 and p5-595 servers.

- 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 setup 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 setup to use a different range of addresses for DHCP
- Additional provision has to be made for HMC connection to the BPC in a powered expansion frame

  If there is a single managed server (with powered expansion frame) then no additional LAN components are required. However, if there are multiple managed servers, additional LAN switch(es) will be needed for the HMC private networks.

The HMC is a pre-installed appliance and is available in a desktop configuration or a rack-mount configuration.

The HMC provides a set of functions that are necessary to manage partition configurations by communicating with the service processor, as follows:

- Logical partitioning control
- Capacity on Demand resource control
- Creation or partition and system profiles
- Boot, start, and stop actions for the system or individual partitions
- Displaying system and partition status

  In a non-partitionable system, the LED codes are displayed in the operator panel. In a partitioned system, the operator panel shows the word LPAR instead of any partition LED codes. Therefore, all LED codes for system partitions are displayed over the HMC.

- An imbedded DVD-RAM for creating and storing configuration backup information
- Cluster support when combined with IBM Cluster Systems Management (CSM) V1.4 or later
- Using a virtual console for each partition or controlled system

  With this feature, every partition can be accessed over the trusted network HMC connection to the server. This is a convenient feature when the partition is not reachable across the public network.
- The HMC provides a Service Focal Point for the systems it controls. It is connected to the service processor of the system using network connection and must be connected to each partition using an Ethernet LAN for Service Focal Point and to coordinate dynamic logical partitioning operations.
- The HMC provides tools for problem determination and service support, such as call-home and error log notification through an analog phone line or Ethernet.

There are several models of HMC that were available for pSeries over the years. These models can be upgraded to manage p5 servers. However, a single HMC cannot manage both POWER4 and POWER5. The upgrade is a reinstall of the HMC code level to one that is compatible with p5 servers.

Whether you opt for a desktop or rack mounted version is personal choice. Clients with space in their rack mounted systems would probably order the rack mounted version with the slide-away keyboard and screen.

- 7310-C04 is a desktop HMC
- 7310-CR3 is rack-mounted HMC

The 7315 models are for the POWER4-based pSeries systems.

For hardware configuration considerations, see Section 5.2.11, “HMC configuration rules” on page 131.

Any 7310 HMC is capable of controlling a p5 system, providing it has the latest software installed on it. It is easy to update your HMC. Keep in mind with the latest software (to support p5 systems) an HMC can no-longer be used to support pSeries POWER4-based servers. You can download HMC updates at:

http://techsupport.services.ibm.com/server/hmc/power5

8.1.1 Desktop HMC

The supported desktop model is the 7310-C04. Older versions can be migrated to POWER5 HMC code level.

On the desktop you can connect a keyboard and mouse to either the standard keyboard, mouse PS/2 style connectors or to the USB ports. A USB keyboard and mouse is part of the ship group for a new HMC.

The desktop HMC can use a number of IBM displays as shown in e-config. A spare display can be used, however there is no ability to add device drivers to the support non-standard displays. Test any display before using it in a production environment.

8.1.2 Rack mounted HMC

The 7310-CR2 can be installed in a 19-inch Rack beside the p5-590 or p5-595 together with the media drawer.

The supported rack mounted models are the 7310-CR3 and the older version 7315-CR2 that can be migrated to POWER5 HMC code. The HMC 7310-CR3 system unit uses a standard 1 EIA unit located below the screen/keyboard mounted 1 EIA pull-out tray.

The rack mount HMC does not have standard PS/2 keyboard and mouse ports. You must order the breakout cable to use them. This cable plugs into the large connector to the left of the RJ-45 connectors. This breakout cable terminates in a pair of PS/2 style female keyboard and mouse connectors and a display connector. You can then plug any standard keyboard or mouse and display.

8.1.3 HMC characteristics

The Hardware Management Console (HMC) controls managed systems, including IBM @server hardware, logical partitions, and Capacity on Demand activation.

The following are the characteristics of the HMC.

- It connects to the managed system or systems using an Ethernet LAN connection.
- It runs a Java-based application running in a pre-installed operating system.
- The user is able to access the management applications through a GUI interface or using a command line interface. Both of these interfaces can be used either by locally operating the HMC (using the keyboard and mouse) or accessing it remotely, through one of the following ways:
  - It can be operated remotely using a Web-based System Management Remote Client running on a Windows based PC, Linux based PC, or AIX workstation.

  - It can also be operated remotely using the command line interface through an SSH connection to the HMC.
- A virtual console terminal can be configured to run on the HMC for each partition reducing the need for extra hardware in each partition.

It is recommended that a second HMC be configured for redundancy. Either HMC can actively manage the same managed systems.

8.2 HMC setup

To successfully configure the HMC, you must understand concepts, make decisions, and prepare information. Depending on the level of customizing you intend to apply to your HMC configuration, you have several options for setting up your HMC to suit your needs.

Depending on the system that is being installed, some of the HMC setup is Customer Set Up (CSU). The IBM Customer Service Representative (CSR) installing the system will perform some of the HMC setup steps and the client will perform the remainder of the steps. It is typical that system partitioning, resource assignment, and other policies be defined by the client.

When a new system or an upgrade is delivered it will have a document named *Start Here* for hardware. This document is the starting point for setting up the system and the HMC. A CD named the IBM ⓔserver Hardware Information Center is included to provide information on the server, the HMC, and other p5 related topics.

The most current version of this information can be accessed directly off the Web. The following is the URL for the US English North American Information Center.

```
http://publib.boulder.ibm.com/infocenter/eserver/v1r2s/en_US/index.htm
```

8.2.1 The HMC logical communications

The following sections describe the communications that are possible.

HMC to managed system

HMC to manage system is a type of communication that is used to perform most of the hardware management functions where the HMC issues control function requests through the service processor of the managed system.

HMC to logical partition

HMC to logical partition is a type of communication that is used to collect platform-related information (hardware error events, hardware inventory) from the operating systems running in the logical partitions, as well as to coordinate platform activities (such as dynamic LPAR and concurrent maintenance) with those operating systems. If you want to use service and error notification features, it is important that you make this connection.

HMC to remote users

HMC to remote users is a type of communication that provides remote users with access to HMC functions. Remote users can access the HMC in the following ways:

- By using the remote client to access all the HMC GUI functions remotely
- By using SSH to access the HMC command line interface remotely
- By using a virtual terminal server for remote access to virtual logical partition consoles

HMC to service provider

HMC to service provided is a type of communication that is used to transmit data, such as hardware error reports, inventory data, and microcode updates, to and from your service provider. You can use this communications path to make automatic service calls.

8.3 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 (BPC) and service processors of the managed systems would be 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 would typically be 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 could do any of the following:
  - An administrative LAN that is separate from the LAN on which all the usual business transactions are running. Remote administrators could 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 (by using a PCI Ethernet card), then that PCI-card becomes the eth0 port. Normally (without the additional card), eth0 is the first of the two integrated Ethernet ports.

8.3.1 Private and open networks in the HMC environment

This topic describes when you might want to use a private network, and when you might want to use an open network.

The connection between the HMC and its managed systems can be implemented either as a open or private network, except with the p5-590 and the p5-595 where the connection must be a private network in a one-HMC configuration.

The term *open* refers to any general, public network that contains elements other than HMCs and service processors that is not isolated behind an HMC. The other network connections on the HMC are considered open, which means that they are configured in a way that you would expect when attaching any standard network device to an open network.

In a *private* network, the only elements on the physical network are the HMC and the service processors of the managed systems. In addition, the HMC provides Dynamic Host Configuration Protocol (DHCP) services on that network that allow it to automatically discover and assign IP configuration parameters to those service processors. You can configure the HMC to select one of several different address ranges to use for this DHCP service, so that the addresses provided to the service processors do not conflict with addresses used on the other networks to which the HMC is connected. The DHCP services allow the elements on the private service network to be automatically configured and detected by the HMC, while at the same time preventing address conflicts on the network.

On a private network all of the elements are controlled and managed by the HMC. The HMC also acts as a functional firewall, isolating that private network from any of the open networks to which the HMC is also attached. The HMC does not allow any IP forwarding; clients on one network interface of the HMC cannot directly access elements on any other network interface.

It is recommended that you implement network communications through a private network, because of the additional security and ease of setup that it provides. However, in some environments, this is not feasible because of physical wiring, hardware availability, floor planning, or control center considerations. In this case, the network communications can be implemented through an open network. The same function is available on both types of networks, although the initial setup and configuration on an open network require more manual steps since the predefined address ranges may be present on your open network.

8.3.2 Using the HMC as a DHCP server

When configuring the HMC as a DHCP server on a private network, you will have the option of selecting from a range of IP addresses for the DHCP server to assign to its clients (the BPCs and service processors of the servers). With p5-590 and p5-595 servers, one HMC must be setup as a DHCP server on the private LAN - this is optional on other p5 systems, where static (or default) IP-addresses could be used. The selectable address ranges include segments from the standard non-routable IP address ranges.

In addition to these standard ranges, there is also a special range of IP addresses that is reserved for this use. This special range can be used to avoid conflicts in cases where the HMC-attached open networks are using one of the non routable address ranges. Based on the range selected, the HMC network interface on the private network will be automatically assigned the first IP address of that range, and the service processors will then be assigned addresses from the rest of the range.

The DHCP server in the HMC uses automatic allocation. This means that each unique service processor Ethernet interface will be reassigned exactly the same IP address each time it is started. Each Ethernet interface has a unique identifier based upon a built-in Media Access Control (MAC) address. This allows the DHCP server to reassign the same IP parameters.

Private direct networking

Private direct networking is an Ethernet connect between port eth0 on the HMC and port HMC 1 on a p5 server (Figure 8-1). On the p5-590 and p5-595 servers, this direct connection would be to the first port of the Bulk Power Controller (this setup is performed during the system installation). For this connection an

Ethernet cross-over is not needed, a standard Ethernet cable will work, because the HMC ports on the service processor are X-ports. The network is provided by the HMC is DHCP serving, DNS, and Firewall. These components can be established by the tasks included in the Guided Setup wizard.

The p5-590 and p5-595 requires one private network connection using DHCP for at least one HMC in the configuration.

The managed system must be powered down and disconnected from any battery or power source prior to completing this operation. Cabling connection from the HMC to the service processor are addressed in Chapter 7.2, “Service processor cabling” on page 172. Detailed information regarding HMC network cabling can also be found in the IBM @server Information Center at the URL listed below; search using the keywords HMC cabling:

```
http://publib.boulder.ibm.com/infocenter/eserver/v1r2s/en_US/index.htm
```

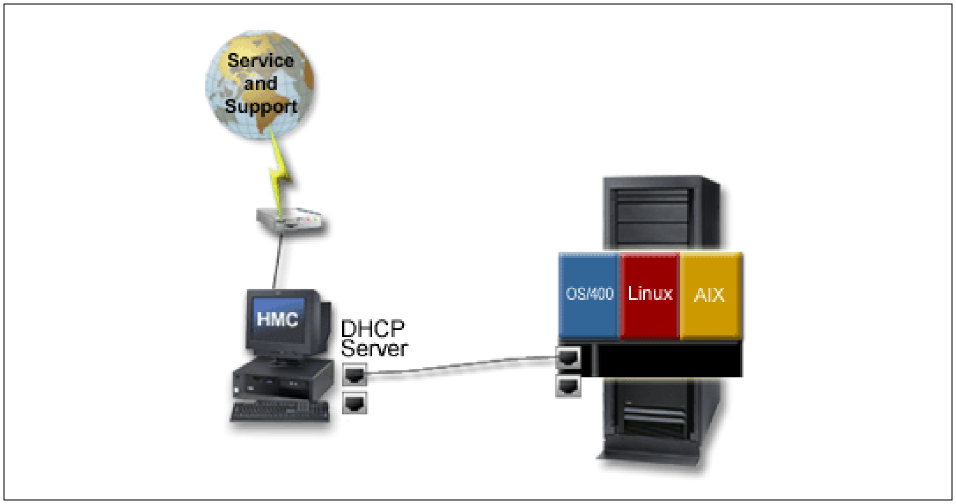


Figure 8-1 Private direct network

Private indirect network

This section describes the general concepts of using an indirect network useful when configuring a second HMC.

A private indirect network (Figure 8-2) is effectively the same as a private network, but the signals pass through one or many hubs or switches. In the graphic below, two servers are connected to a hub and two HMCs connected to the same hub. One HMC is connected to IBM Service and Support. This HMC is actively managing both the servers. The second HMC is connected but redundant, in that it is not actively managing either of the servers. We do not have any recommendation for the type of hub or switch that can be used in this arrangement other than it supports industry standard Ethernet.

We suggest that you connect the HMC directly to the server before installing if on an active network so you are confident with the HMC operation and management.

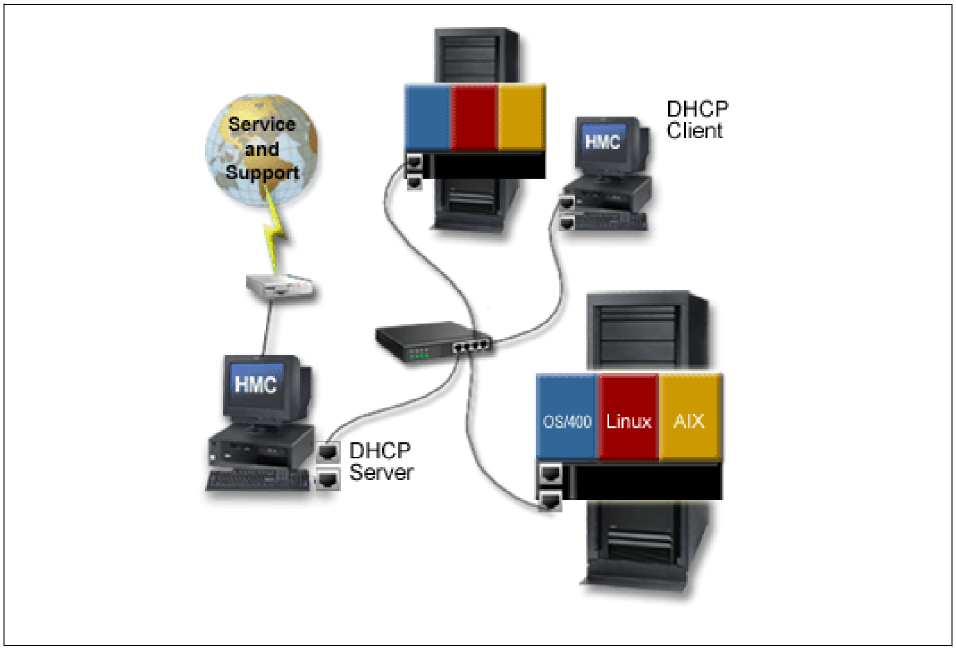


Figure 8-2 HMC with hub/switch attachment

Private network and a public network

As there are two network connections on an HMC you can connect it to both a private network and a public network. The HMC (Figure 8-3) is the DHCP server to the indirect connected servers on the private network. There is also a connection to the local public network as a DHCP client. in this network we have Web-based System Management Remote Clients (WSMRC) and a remote HMC. This arrangement is probably what most large clients would choose. The *main* HMC is located in the dark machine room. The remote HMC and could be located in the bridge or IT department for operational use.

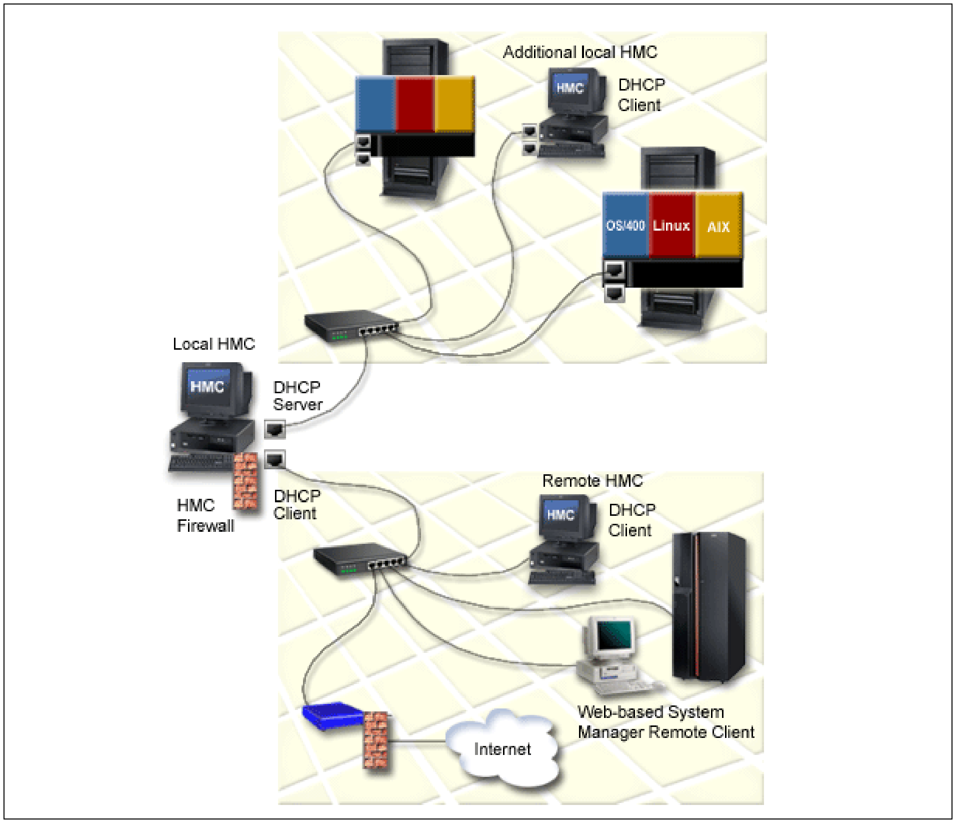


Figure 8-3 HMC attached to both private and public network

By default, the default DHCP IP address range is 192.168.0.2 - 192.168.255.254. The HMCs network address is 192.168.0.1 and the network mask is 255.255.0.0.

Note: With a desktop HMC, by default, only one Ethernet port is provided. To provide the private and open network connections, an Ethernet adapter is required. In this case, eth0 remains on the integrated port.

8.3.3 HMC connections

For the p5-590 and p5-595 servers, an HMC must be directly attached. In general, we recommend private service networks for all systems because of the simplified setup and greater security, however a private network is required between HMC and the p5-590 and p5-595 servers because of the following reasons:

- The Bulk Power Controllers are dependent upon the HMC to provide them with DHCP addresses (there is no way to set a static addresses)
- The HMC must distribute information to the SPs and BPCs about the topology of the network.
- The SPs acquire their TCP/IP address information from the HMC as they are connected to the HMC through an internal hub in the BPC, and then through the network connection from the BPC to the HMC.

The use of two HMCs requires the use of different DHCP address ranges, so that the two Ethernet interfaces in each service processor and BPC are configured on different subnets.

If there are multiple systems (additional Managed Server Frames) then an external hub or switch is required for each of the private networks.

Figure 8-4 shows the connections between a primary and secondary HMC and the redundant BPCs in a system frame. If additional powered frames are attached, hubs are used to connect the network. Notice the need for a network connection between each HMC and the corresponding BPC in a given configuration.

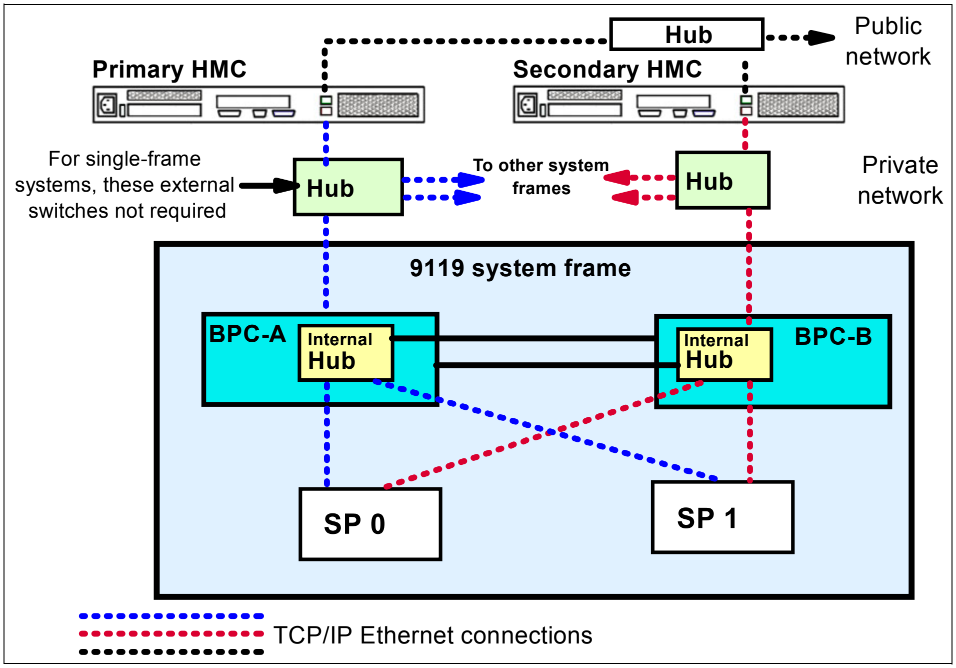


Figure 8-4 Primary and secondary HMC to BPC connections

8.3.4 Predefined HMC user accounts

The HMC users (login accounts) are associated with specific definitions of authorized tasks and resources (roles). The default (predefined) roles and users are as follows:

super administrator The super administrator acts as the root user, or *manager* of the HMC system. The super administrator has unrestricted authority to access and modify most of the HMC system.

hscroot user A product engineer that assists in support situations, but cannot access HMC user management or partition functions. To provide support access for your system, you must create and administer a user ID of hscpe with the hmcpe (product engineer) role.

service representative A service representative is an employee who is at your location to install, configure, or repair the system.

operator An operator is responsible for daily system operation.

product engineer A product engineer assists in support situations, but cannot access HMC user management functions. To provide support access for your system, you must create and administer user IDs with the product engineer role.

viewer A viewer can view HMC information, but cannot change any configuration information.

The HMC predefined users include invscout, which is used internally for the collection of VPD (Vital Product Data). This is not a login account, but must not be deleted.

Pre-defined passwords for hscroot and root are included with the HMC. It is imperative to your systems security that you change all predefined passwords during the initial setup of the system (Table 8-1). The root user is not accessible except at boot time for file system check and recovery operations.

Table 8-1 HMC user passwords

User	Password	Purpose
hscroot	abc123	Is used to login for the first time.
root	passw0rd	Only for authorized service provider

The Advanced System Management Interface (ASMI) is a service that is provided to allow access to the system to perform actions including powering on the system, changing the boot configuration, and other tasks. ASMI can be launched from a web browser or from the HMC. ASMI access methods, services, and function are addressed in Section 7.3.2, “Accessing ASMI using a Web browser” on page 177.

8.4 HMC login

After a normal power on of the HMC, the HMC and login as user hscroot and the default password of abc123, Figure 8-5 appears.

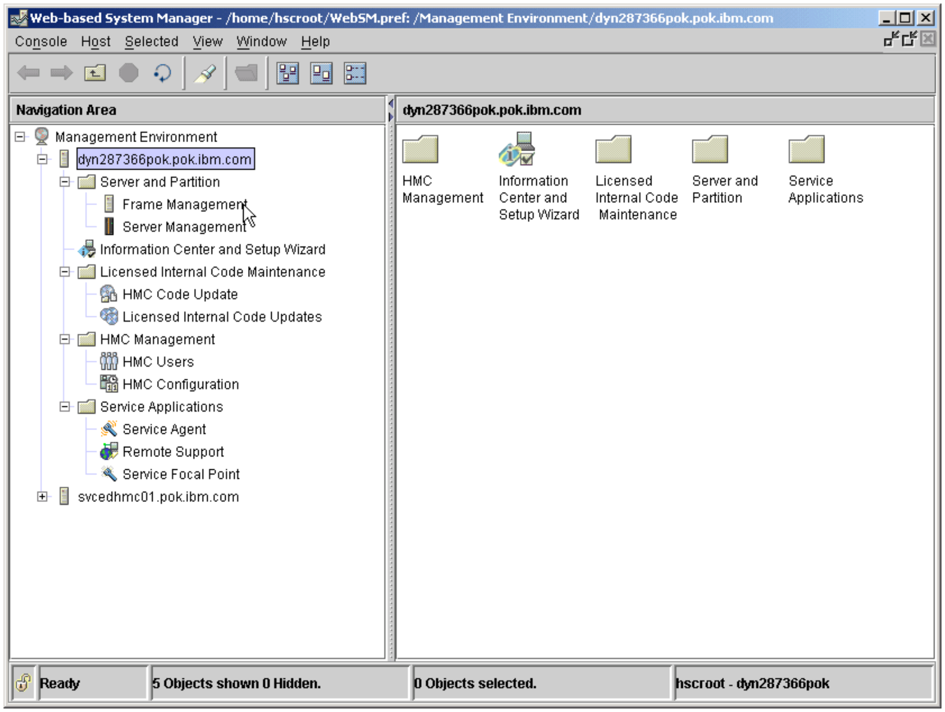


Figure 8-5 First screen after login as hscroot user

8.4.1 Required setup information

The IBM @server Hardware Information Center guides you through cabling up the HMC and then through a checklist to gather information needed to configure the HMC. The information needed includes:

- HMC hostname
- Domain name
- Enabling DHCP server support on the HMC

- Enabling DHCP client support on the HMC
- A gateway IP address
- Firewall settings
- Enabling DNS
- Client contact information as the HMC will be a service focal point
- Type of connection to connect to the service provider

8.5 HMC Guided Setup Wizard

The Guided Setup wizard is a tool on the HMC designed to make the setup of the HMC quick and easy. The Guided Setup Wizard can be run only on the HMC itself, it cannot be run using the Web-based System Management Remote Client.; you can choose a fast path through the wizard to quickly create the recommended HMC environment, or you can choose to fully explore the available settings that the wizard guides you through. You can also perform the configuration steps without the aid of the wizard.

Both the desktop and rack mounted HMC models come with the pre-installed operating system, HMC code, browser, and IBM eserver Hardware Information Center pre-loaded. Therefore, the HMC can be unpacked and powered up immediately after the networking components, power, and other basic connections are made.

Note: Refer to the setup instructions that came with the HMC. Before connecting the power to the HMC, it is essential to check the setting for line-voltage. On the back of the system, near the power cord receptacle, there is a dual-range switch (110/240v) that must be at the correct setting.

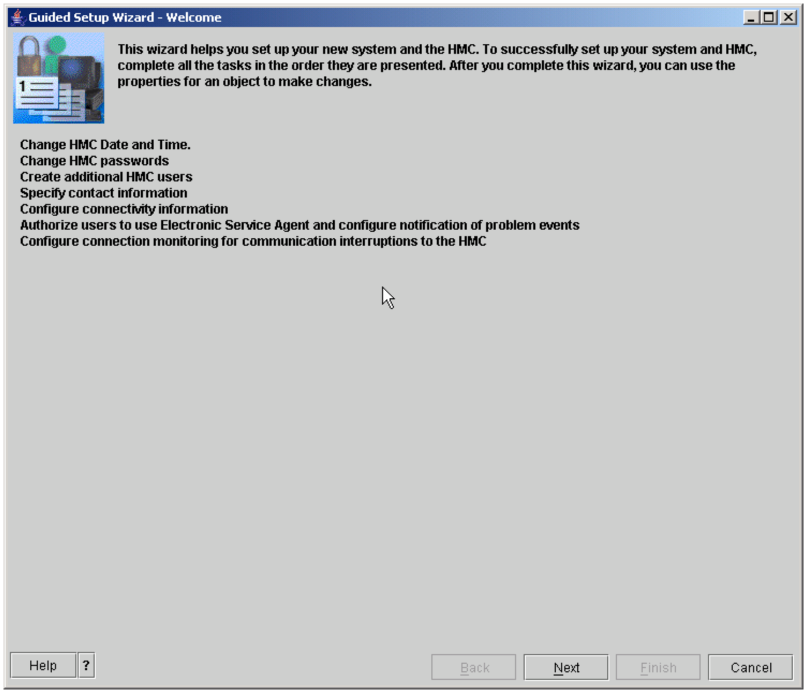
The Guided Setup wizard guides you through the main tasks needed to configure the HMC in very logical steps. To access the setup wizard and information center, power on the HMC.

Note: When installing a new p5 system, do not power on the system before connecting it to an HMC. The service processor on a p5 system is a DHCP client and will search for a DHCP server (the HMC) to obtain its IP address. If no DHCP server can be found, the service processor will assign a default IP address and simple communication with the HMC will be prevented. If this occurs you will have to manually change the IP setting of the service processor.

The wizard will launch automatically the first time the HMC is started. Using this wizard is the simplest way to configure your HMC. Before using the Guided Setup wizard you must understand the main concepts of HMC and decide which functions are relevant to your environment. The Guided Setup will enable you to configure the following functions:

- Set date and time
- Change the passwords of the predefined user ids for hscroot and root
- Create new user IDs
- Configure network settings
- Client contact information for support and services
- Configure connectivity for service-related activities
- Configure Service Focal Point

The Guided Setup Wizard appears as in Figure 8-6.

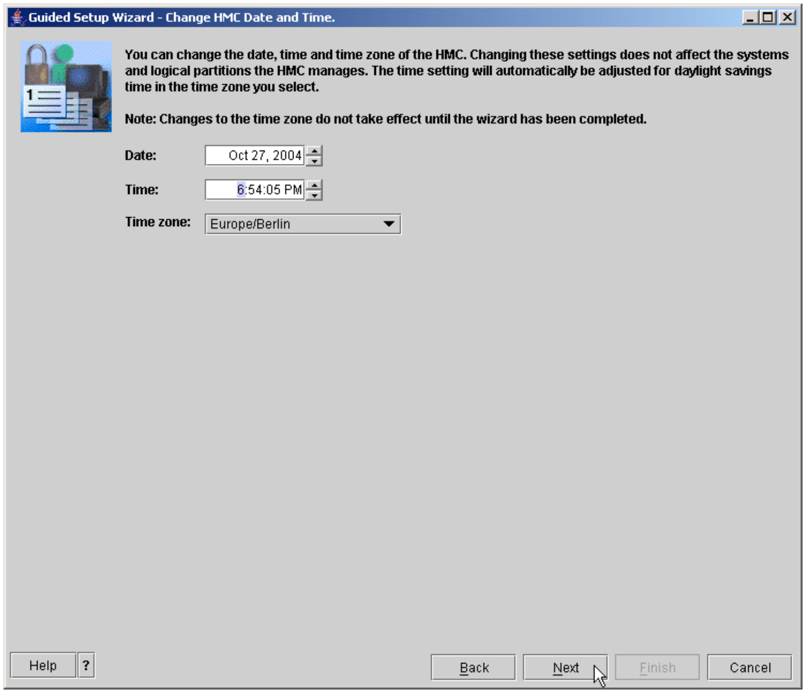


This wizard helps you set up your new system and the HMC. To successfully set up your system and HMC, complete all the tasks in the order they are presented. After you complete this wizard, you can use the properties for an object to make changes.

Change HMC Date and Time.
Change HMC passwords
Create additional HMC users
Specify contact information
Configure connectivity information
Authorize users to use Electronic Service Agent and configure notification of problem events
Configure connection monitoring for communication interruptions to the HMC

Figure 8-6 Guided Setup Wizard

On the Guided Setup wizard - Change HMC Data and Time screen (Figure 8-7), enter the correct date, time, and time zone for your environment. This is typically the time zone of the server, assuming the HMC is local to the machine. For remote machines you must decide which is the correct time zone for your environment.



Guided Setup Wizard - Change HMC Date and Time.

You can change the date, time and time zone of the HMC. Changing these settings does not affect the systems and logical partitions the HMC manages. The time setting will automatically be adjusted for daylight savings time in the time zone you select.

Note: Changes to the time zone do not take effect until the wizard has been completed.

Date: Oct 27, 2004

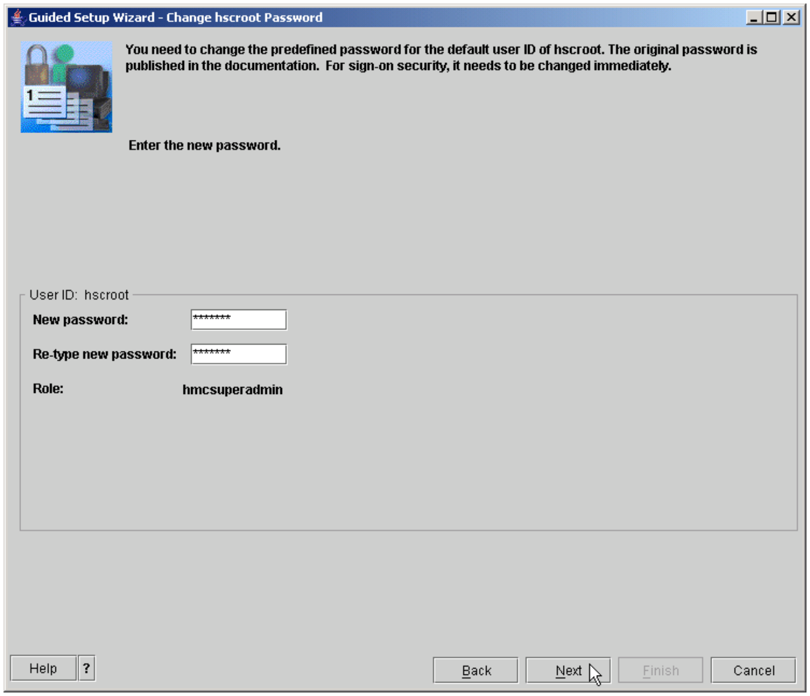
Time: 6:54:05 PM

Time zone: Europe/Berlin

Figure 8-7 Date and Time settings

The Guided Setup Wizard - Change hscroot Password screen is now displayed as shown in (Figure 8-8). Enter the current hscroot password (normally this should be the default password of abc123) and then the new password you would like.

> **Note:** There are specific restrictions regarding acceptable characters used in passwords on the HMC. For example, characters must be standard 7-bit ASCII. For more information, click on the Help button at the bottom of the HMC dialog.



Guided Setup Wizard - Change hscroot Password

You need to change the predefined password for the default user ID of hscroot. The original password is published in the documentation. For sign-on security, it needs to be changed immediately.

Enter the new password.

User ID: hscroot

New password: *******

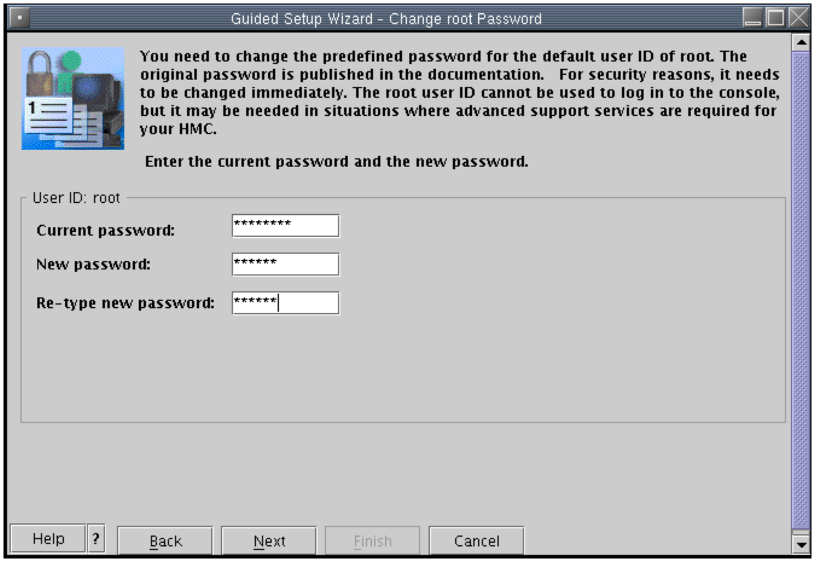
Re-type new password: *******

Role: hmcsuperadmin

Figure 8-8 The hscroot password

The Change root Password screen is now displayed as shown in (Figure 8-9).

The root user ID is used by the authorized service provider to perform maintenance procedures and can not be used to directly log in to the HMC. Enter the current root password, normally this should be the default password of passw0rd - where 0 is the number zero rather than the letter o. Enter the new password (minimum 7 characters) you would like for the root user ID. The root user is need for maintenance during boot of the HMC such as **`fsck`**.



Guided Setup Wizard - Change root Password

You need to change the predefined password for the default user ID of root. The original password is published in the documentation. For security reasons, it needs to be changed immediately. The root user ID cannot be used to log in to the console, but it may be needed in situations where advanced support services are required for your HMC.

Enter the current password and the new password.

User ID: root

Current password: ********

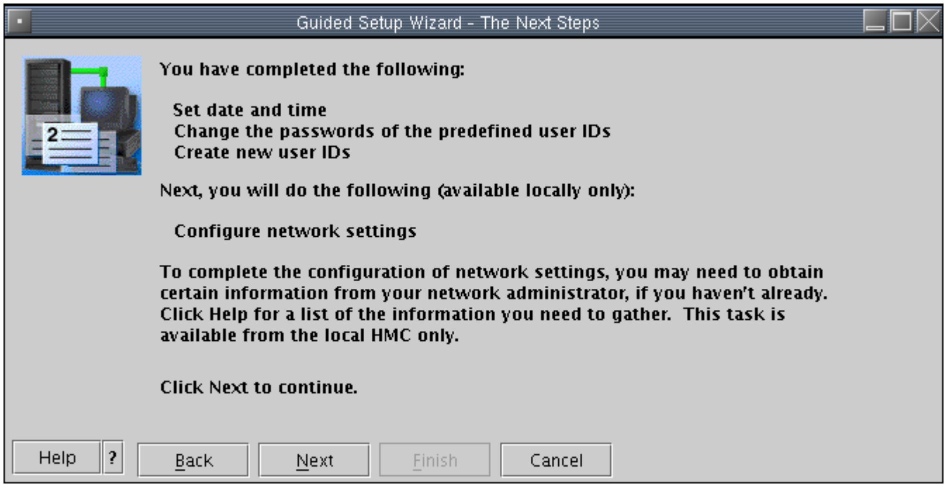
New password: ******

Re-type new password: ******

Figure 8-9 The root password

The Create additional HMC users screen will be shown next (no figure is available for this dialog). You can now optionally create new HMC users at this stage. In our example, we decided to create a new hscpe user ID with a role of hmcpe to allow support staff limited access to HMC (for maintenance).

This completes the first part and you see the next screen (Figure 8-10).



Guided Setup Wizard - The Next Steps

You have completed the following:

Set date and time
Change the passwords of the predefined user IDs
Create new user IDs

Next, you will do the following (available locally only):

Configure network settings

To complete the configuration of network settings, you may need to obtain certain information from your network administrator, if you haven't already. Click Help for a list of the information you need to gather. This task is available from the local HMC only.

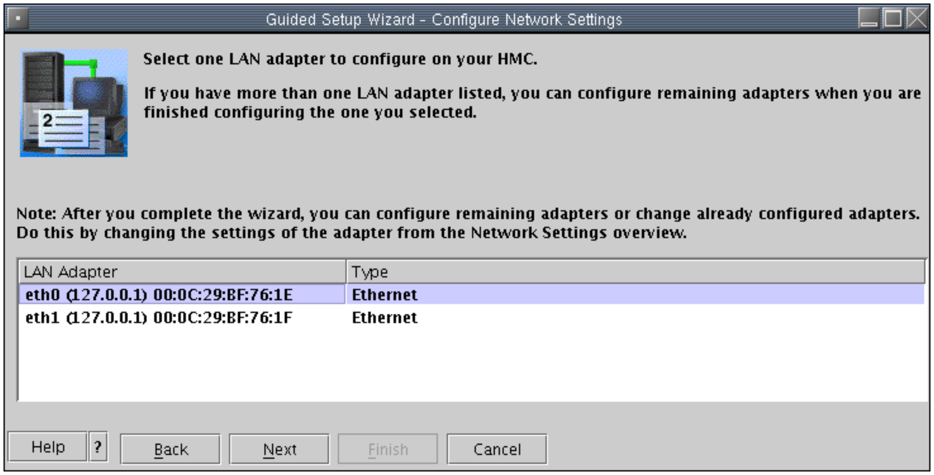
Click Next to continue.

Figure 8-10 First part of Guided Setup Wizard is done

The next section will configure the HMC network settings. You will need to have planned your network environment for HMC before continuing with these next tasks. You should use the values entered in the HMC Guided Setup wizard checklist.In our example we show how to configure the HMC for both a private and open network.

Use the first Ethernet network card (eth0) in the HMC for the private network configuration. Note that the Ethernet ports are called eth0 and eth1 on the HMC. On a p5 managed system the HMC ports are labeled HMC1 and HMC2 (or Port A and Port B). For the p5-590 and p5-595, the HMC connects to the BPC.

Select one LAN adapter for configuration (Figure 8-11).



Select one LAN adapter to configure on your HMC.

If you have more than one LAN adapter listed, you can configure remaining adapters when you are finished configuring the one you selected.

Note: After you complete the wizard, you can configure remaining adapters or change already configured adapters. Do this by changing the settings of the adapter from the Network Settings overview.

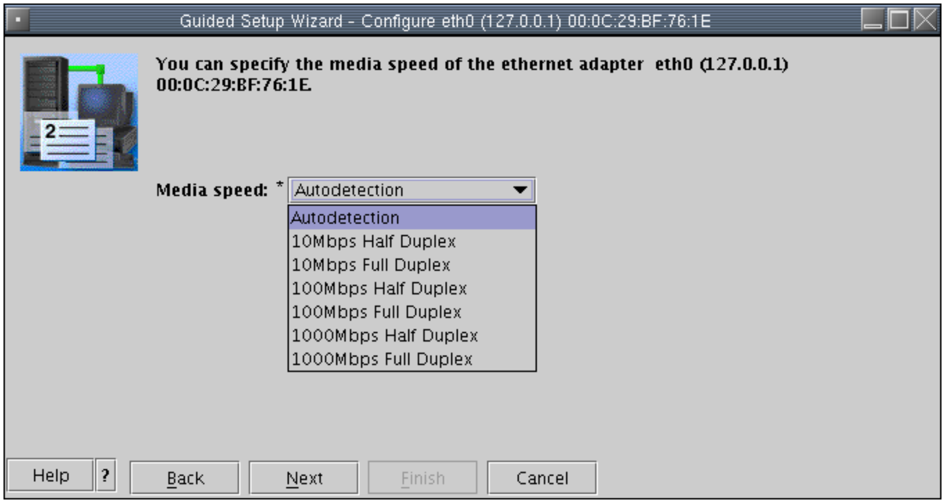
LAN Adapter	Type
eth0 (127.0.0.1) 00:0C:29:BF:76:1E	Ethernet
eth1 (127.0.0.1) 00:0C:29:BF:76:1F	Ethernet

Figure 8-11 Select LAN adapter

If you only see one Ethernet adapter, than you must enable the second adapter in the bios setup.

> **Note:** If Ethernet ports in the HMC are not active, check the SETUP of the ports:
>
> - Reboot HMC
> - Press F1 to go to the BIOS setup screen
> - In the IBM Setup Utility, check Network devices are set to enable

You can normally leave the LAN adapter speed at automatic detection for the initial setup. This is the best selection for the private service network connection. But if you use a private indirect connection over a client Ethernet switch or hub, media speed may need to be set (Figure 8-12).



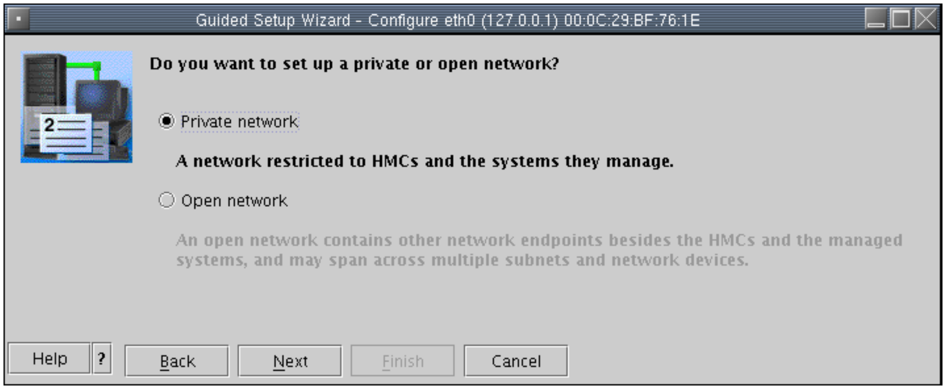
Guided Setup Wizard - Configure eth0 (127.0.0.1) 00:0C:29:BF:76:1E

You can specify the media speed of the ethernet adapter eth0 (127.0.0.1) 00:0C:29:BF:76:1E.

Media speed: * Autodetection

Figure 8-12 Speed selection

The next panel allows you to specify whether network is private or open (Figure 8-13). Ethernet 0 is normally used for a private service network all other network adapter are open network. A private network connection using DHCP is required on the p5-590 and p5-595.



Guided Setup Wizard - Configure eth0 (127.0.0.1) 00:0C:29:BF:76:1E

Do you want to set up a private or open network?

● Private network

A network restricted to HMCs and the systems they manage.

○ Open network

An open network contains other network endpoints besides the HMCs and the managed systems, and may span across multiple subnets and network devices.

Figure 8-13 Network type

In the example, we choose the private service network. Go to the DHCP configuration. The HMC provides Dynamic Host Configuration Protocol (DHCP) services to all clients in a private network. These clients will be our managed systems and other HMCs. You can configure the HMC to select one of several different IP address ranges to use for this DHCP service so that the addresses provided to the managed systems do not conflict with addresses used on other networks that the HMC is connected. We have a choice of standard non routable IP address ranges that will be assigned to its clients. The ranges we can select from are:

- 192.168.0.2 - 192.168.255.254
- 172.16.0.3 - 172.16.255.254
- 172.17.0.3 - 172.17255.254
- 10.0.0.2 10.0.0.254
- 10.0.128.2 - 10.0.143.254
- 10.0.255.2 - 10.0.255.254
- 10.1.0.2 - 10.1.15.254
- 10.1.255.2 - 10.1.255.254
- 10.127.0.2 - 10.127.15.254
- 10.127.255.2 - 10.127.255.254
- 10.128.0.2 - 10.128.15.254
- 10.128.128.2 - 10.128.128.254
- 10.128.240.2 - 10.128.255.254
- 10.254.0.2 - 10.254.0.254
- 10.254.240.2 - 10.254.255.254
- 10.255.0.2 - 10.255.0.254
- 10.255.128.2 - 10.255.143.254
- 10.255.255.2 - 10.255.255.254
- 9.6.24.2 - 9.6.24.254
- 9.6.25.2 - 9.6.25.254

The 9.6.24.2 - 9.6.25.254 range is a special range of IP address that can be used to avoid conflicts in cases where the HMC is attached an open network that is already using non routable addresses.

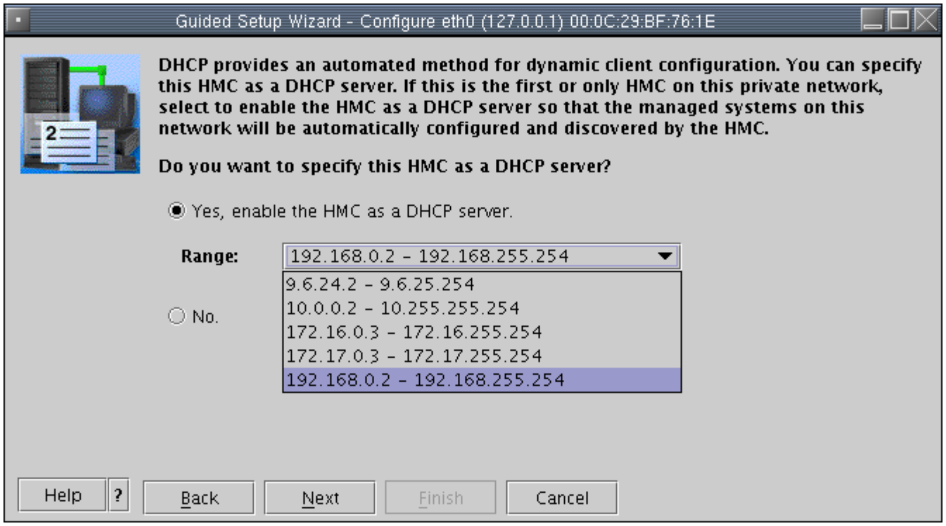
The HMC LAN adapter will be assigned the first IP address out of the range selected. In our example, we select the 192.168.0.2 - 192.168.255.254 range so our HMC (Figure 8-14) will be given IP address 192.168.0.1. Any other client (HMC or managed system) will also be given an address from this range.

The DHCP server uses each client's built in Media Access Control (MAC) address to ensure that it will reassign each client with same IP address as before. When a managed system starts it will try and contact the DHCP service to obtain its IP address.

Note: The information about IP address and MAC is part of the Backup Critical data task that you should do regularly on the HMC. If you ever need to recover the HMC (re-install it), the data saved on the backup will be restored including the LAN configuration of the HMC.

If the managed system is unable to contact the HMC DHCP service then, the managed system will use its last given IP address.

If this is the first (or only) HMC connected to the managed servers, select the 192.168.0.2 - 192.168.255.254 range and click Next to continue. If this is the secondary (or redundant) HMC connected to the same p5-590 or p5-595 servers, then you must select a range of addresses which is different from the range which was selected on the first HMC.


Guided Setup Wizard - Configure eth0 (127.0.0.1) 00:0C:29:BF:76:1E

DHCP provides an automated method for dynamic client configuration. You can specify this HMC as a DHCP server. If this is the first or only HMC on this private network, select to enable the HMC as a DHCP server so that the managed systems on this network will be automatically configured and discovered by the HMC.

Do you want to specify this HMC as a DHCP server?

● Yes, enable the HMC as a DHCP server.

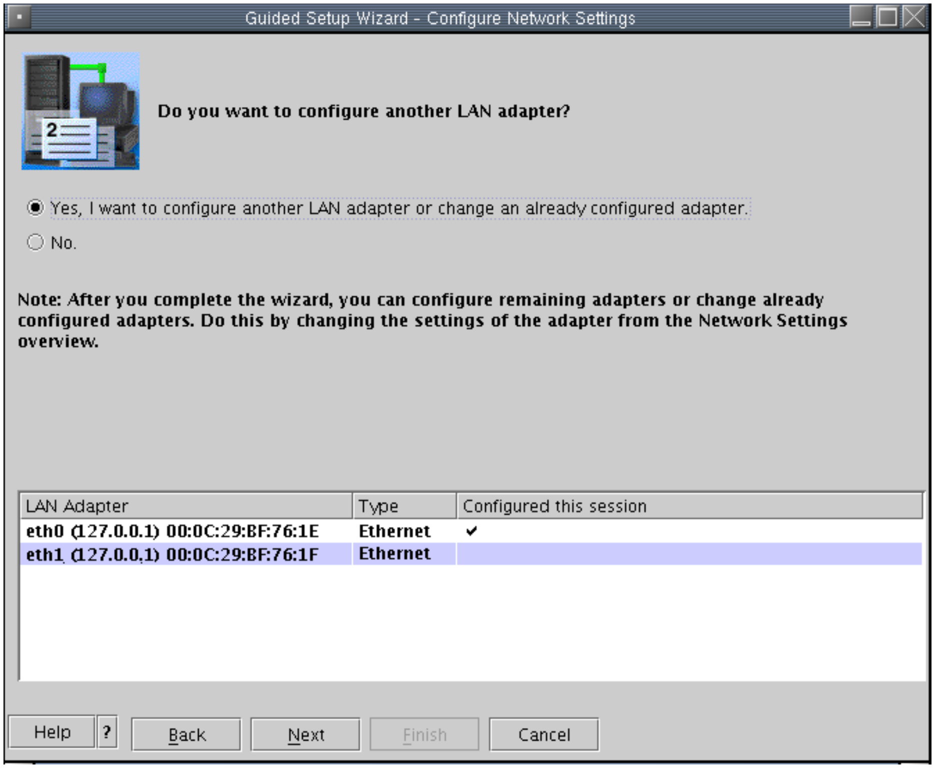
Range: 192.168.0.2 - 192.168.255.254

9.6.24.2 - 9.6.25.254
10.0.0.2 - 10.255.255.254
172.16.0.3 - 172.16.255.254
172.17.0.3 - 172.17.255.254
192.168.0.2 - 192.168.255.254

○ No.

Figure 8-14 Configure eth0 DHCP range

If you select YES, you will be asked to configure the next adapter, with NO you are sent to the next configuration point, the HMC host name and domain. If you configure a second LAN adapter, you must also configure a firewall. This will be discussed later on. For this example, select NO (Figure 8-15).



Do you want to configure another LAN adapter?

● Yes, I want to configure another LAN adapter or change an already configured adapter.

○ No.

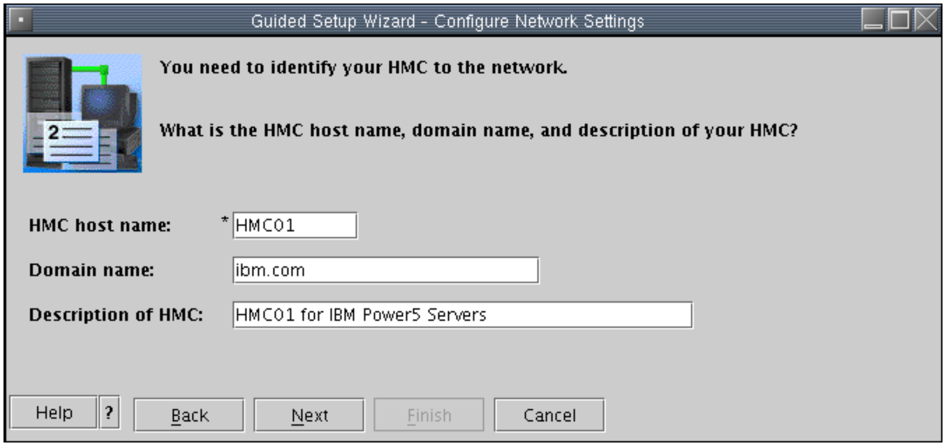
Note: After you complete the wizard, you can configure remaining adapters or change already configured adapters. Do this by changing the settings of the adapter from the Network Settings overview.

LAN Adapter	Type	Configured this session
eth0 (127.0.0.1) 00:0C:29:BF:76:1E	Ethernet	✓
eth1 (127.0.0.1) 00:0C:29:BF:76:1F	Ethernet	

Figure 8-15 Second LAN adapter

Define the host name. The domain name is optional.

Note: The hostname will be used in TCP/IP communications between the HMC and the LPARs on the public network. It is, therefore, important to ensure a consistent naming convention: all LPARs, and the HMC, must use either short-names (hostname only) or long-names (fully qualified domain name) plus the domain name. If you are using DNS, then a fully qualified domain name must be used.



You need to identify your HMC to the network.

What is the HMC host name, domain name, and description of your HMC?

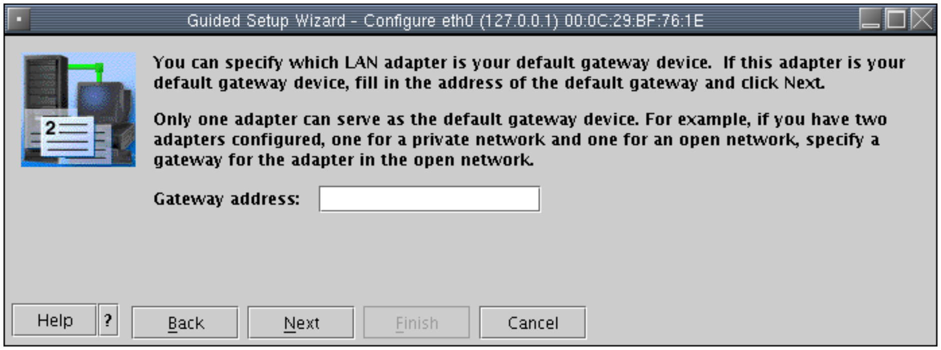
HMC host name: * HMC01

Domain name: ibm.com

Description of HMC: HMC01 for IBM Power5 Servers

Figure 8-16 Host name and domain name

Next, the default gateway is assigned. In our case we have a private network with direct connection to the service processor and we do not need a default gateway (Figure 8-17).


Guided Setup Wizard - Configure eth0 (127.0.0.1) 00:0C:29:BF:76:1E
2
Help
?
Back
Next
Finish
Cancel

You can specify which LAN adapter is your default gateway device. If this adapter is your default gateway device, fill in the address of the default gateway and click Next.

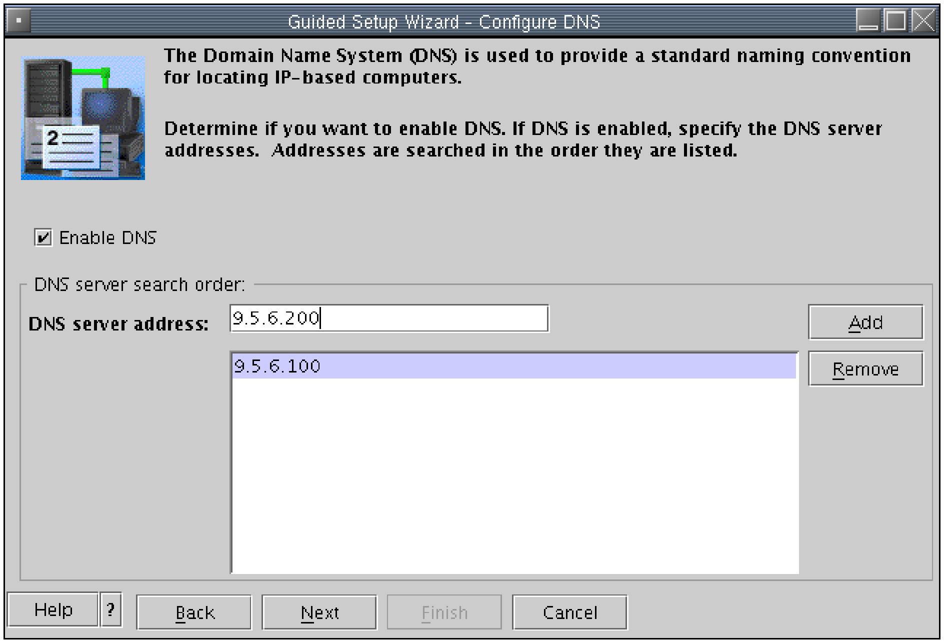
Only one adapter can serve as the default gateway device. For example, if you have two adapters configured, one for a private network and one for an open network, specify a gateway for the adapter in the open network.

Gateway address:

Figure 8-17 Default gateway IP address

If we have an open network, you must define a default gateway. This is a requirement for DLAR functions.

If you have an open network, most clients will also setup a DNS configuration. This is the DNS on the open network using eth1 of the HMC. In our private network, this isn’t required. Press the Next button to move ahead (Figure 8-18).



Guided Setup Wizard - Configure DNS

The Domain Name System (DNS) is used to provide a standard naming convention for locating IP-based computers.

Determine if you want to enable DNS. If DNS is enabled, specify the DNS server addresses. Addresses are searched in the order they are listed.

☑ Enable DNS

DNS server search order:

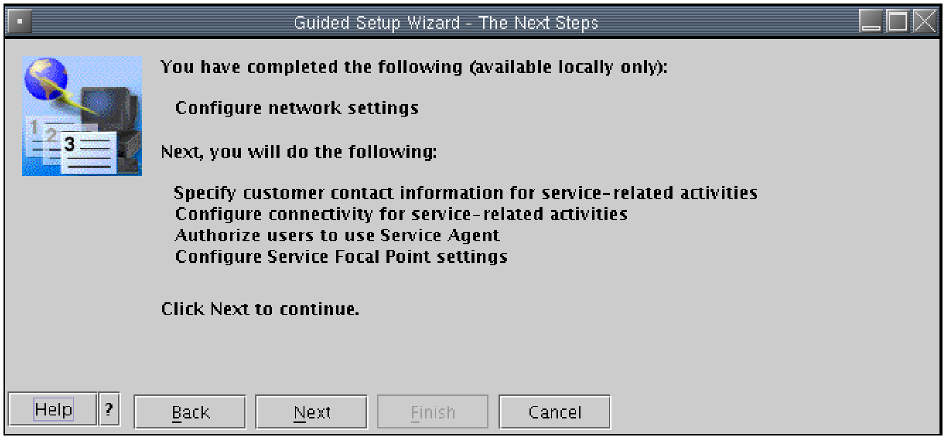
DNS server address: 9.5.6.200

9.5.6.100

Figure 8-18 DNS IP address

The panel about the configuration of a firewall is next. The client can set up the firewall settings. This is not needed for a basic installation.

This ends the IP configuration in the Setup Wizard (Figure 8-19).



Guided Setup Wizard - The Next Steps

You have completed the following (available locally only):

Configure network settings

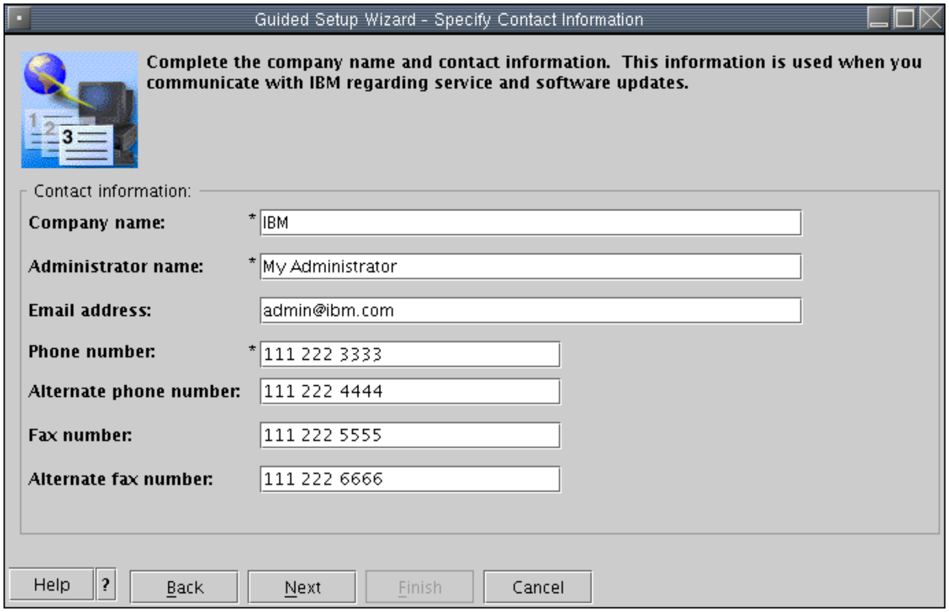
Next, you will do the following:

Specify customer contact information for service-related activities
Configure connectivity for service-related activities
Authorize users to use Service Agent
Configure Service Focal Point settings

Click Next to continue.

Figure 8-19 End of network configuration

Next, you need the client data for the contact information. This information will be included in the problem record (Figure 8-20 and Figure 8-21). If a problem record is called home, the information entered through this dialog is used.



Complete the company name and contact information. This information is used when you communicate with IBM regarding service and software updates.

Contact information:

Company name: * IBM

Administrator name: * My Administrator

Email address: admin@ibm.com

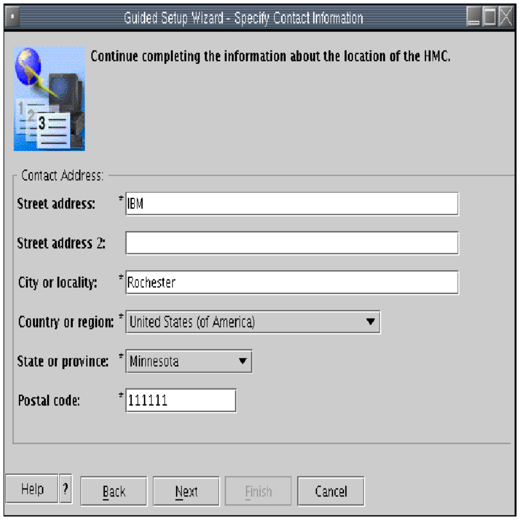
Phone number: * 111 222 3333

Alternate phone number: 111 222 4444

Fax number: 111 222 5555

Alternate fax number: 111 222 6666

Figure 8-20 Client contact information



Continue completing the information about the location of the HMC.

Contact Address:

Street address: * IBM

Street address 2:

City or locality: * Rochester

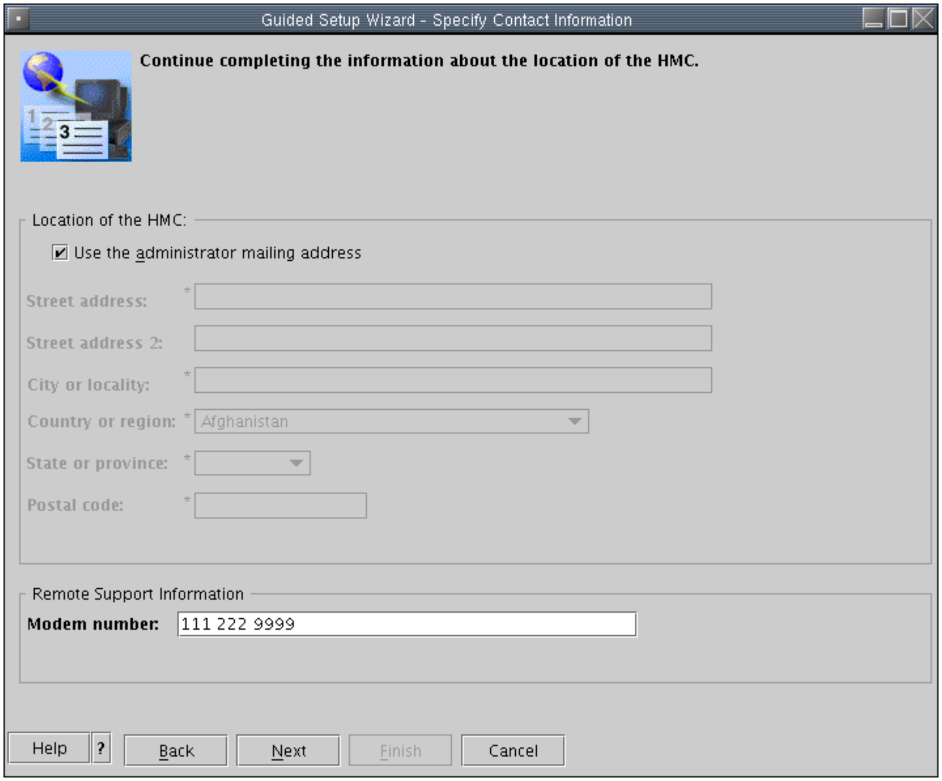
Country or region: * United States (of America) ▼

State or province: * Minnesota ▼

Postal code: * 111111

Figure 8-21 Client contact information

The last screen for the Contact Information is shown next. You should enter the location details of this HMC and the p5-590 or p5-595. If the location address is the same as the contact address used in the previous step, then click Use the administrator mailing address. Otherwise fill in the correct HMC location address details (Figure 8-22).



Guided Setup Wizard - Specify Contact Information

Continue completing the information about the location of the HMC.

Location of the HMC:

☑ Use the administrator mailing address

Street address: *

Street address 2:

City or locality: *

Country or region: * Afghanistan

State or province: *

Postal code: *

Remote Support Information

Modem number: 111 222 9999

Figure 8-22 Remote support information

The Guided Setup Wizard - Configure Connectivity to Your Service Provider screen allows you to select the desire communications method used to connect to IBM for service and support related functions. There are four service applications available on the HMC:-

- **Electronic Service Agent**

  Monitors your managed systems for problems and if enabled, reports them electronically to IBM.

- **Inventory Scout Services**

  Reports hardware and software information from your managed systems to IBM.

- **Remote Support Facility**

  Enables the HMC to call out to IBM for problem reporting as well as enabling remote access to your managed systems (if enabled by a client).

- **Service Focal Point**

  Collects system management issues in one central point for all your partitions.

You can select the connectivity method you wish to use when communicating electronically with IBM. There are three options available to you:

- Dial-up from the local HMC

  This option will use the IBM supplied modem with the HMC to dial in to the IBM support network. You may choose this option if your HMC doesn’t have a high speed Internet connection through an open network or has only been configured in a private network.

- Virtual private network (VPN) through the Internet

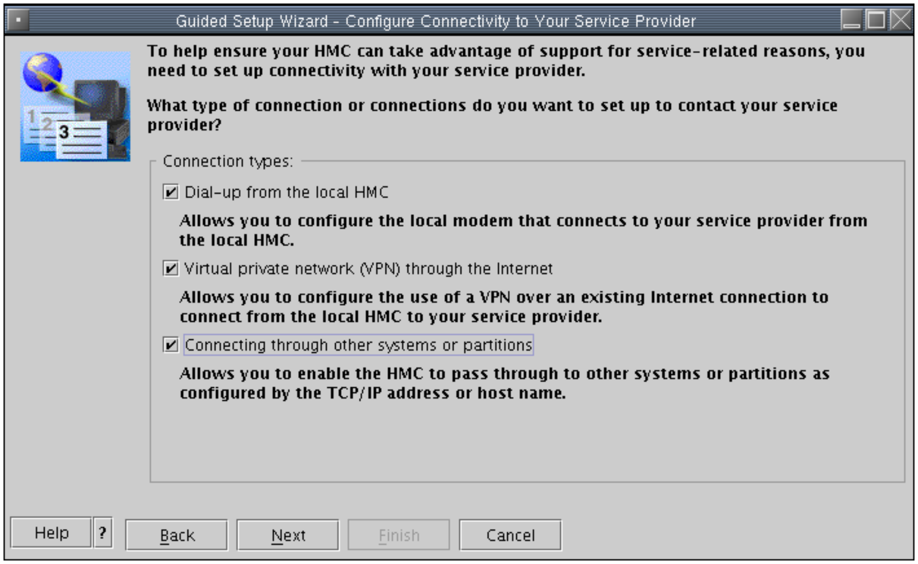
  This option will use a high speed Internet connection to connect to the IBM support network. This is the fastest option available on the HMC, however your company may restrict this type of connection.

  The minimum requirements for call home service connection through a firewall requires port 500 (TCP and UDP) and port 4500 (UDP). If the HMC is located behind a NAT router, end points 207.25.252.196 and 129.42.160.16 should be specified.

- Connecting through other systems or partitions over the intranet

  This option sends information through another system in your network that can connect to IBM. The pass-through system could be another HMC (only supported passthrough servers are i5/OS based systems).

Next is the configuration for call home with the modem (Figure 8-23).


Guided Setup Wizard - Configure Connectivity to Your Service Provider

To help ensure your HMC can take advantage of support for service-related reasons, you need to set up connectivity with your service provider.

What type of connection or connections do you want to set up to contact your service provider?

Connection types:

☑ Dial-up from the local HMC

Allows you to configure the local modem that connects to your service provider from the local HMC.

☑ Virtual private network (VPN) through the Internet

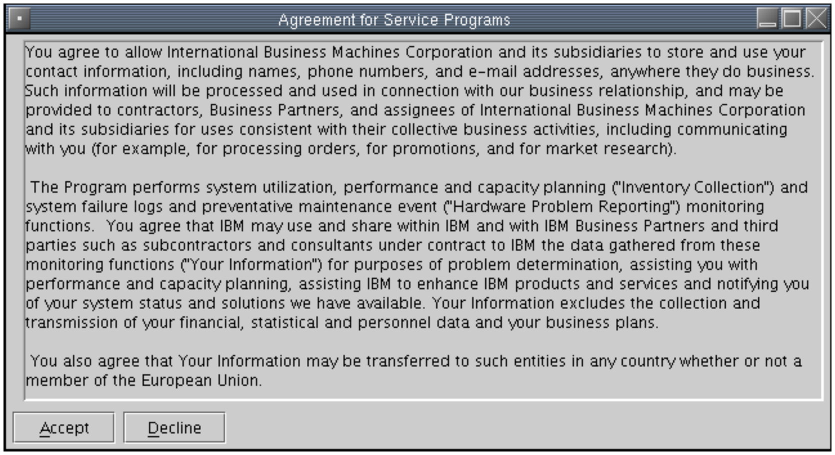
Allows you to configure the use of a VPN over an existing Internet connection to connect from the local HMC to your service provider.

☑ Connecting through other systems or partitions

Allows you to enable the HMC to pass through to other systems or partitions as configured by the TCP/IP address or host name.

Figure 8-23 Callhome connection type

Next screen is the Agreement for Service Programs and you must review and accept in order to continue (Figure 8-24).



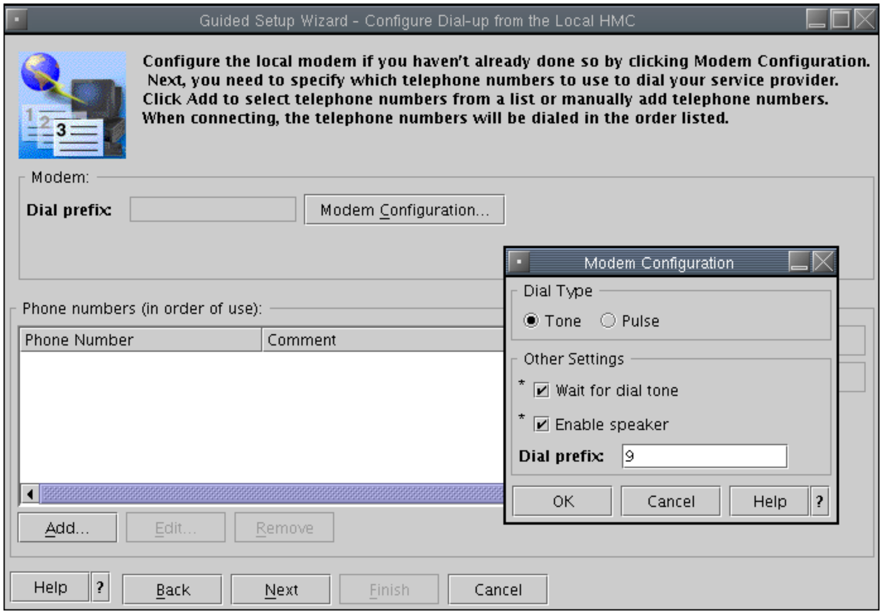
You agree to allow International Business Machines Corporation and its subsidiaries to store and use your contact information, including names, phone numbers, and e-mail addresses, anywhere they do business. Such information will be processed and used in connection with our business relationship, and may be provided to contractors, Business Partners, and assignees of International Business Machines Corporation and its subsidiaries for uses consistent with their collective business activities, including communicating with you (for example, for processing orders, for promotions, and for market research).

The Program performs system utilization, performance and capacity planning ("Inventory Collection") and system failure logs and preventative maintenance event ("Hardware Problem Reporting") monitoring functions. You agree that IBM may use and share within IBM and with IBM Business Partners and third parties such as subcontractors and consultants under contract to IBM the data gathered from these monitoring functions ("Your Information") for purposes of problem determination, assisting you with performance and capacity planning, assisting IBM to enhance IBM products and services and notifying you of your system status and solutions we have available. Your Information excludes the collection and transmission of your financial, statistical and personnel data and your business plans.

You also agree that Your Information may be transferred to such entities in any country whether or not a member of the European Union.

Figure 8-24 Licence agreement

Here you must define the modem configuration (Figure 8-25).



Guided Setup Wizard - Configure Dial-up from the Local HMC

Configure the local modem if you haven't already done so by clicking Modem Configuration. Next, you need to specify which telephone numbers to use to dial your service provider. Click Add to select telephone numbers from a list or manually add telephone numbers. When connecting, the telephone numbers will be dialed in the order listed.

Modem:

Dial prefix:

Phone numbers (in order of use):

Phone Number	Comment

Modem Configuration

Dial Type

● Tone ○ Pulse

Other Settings

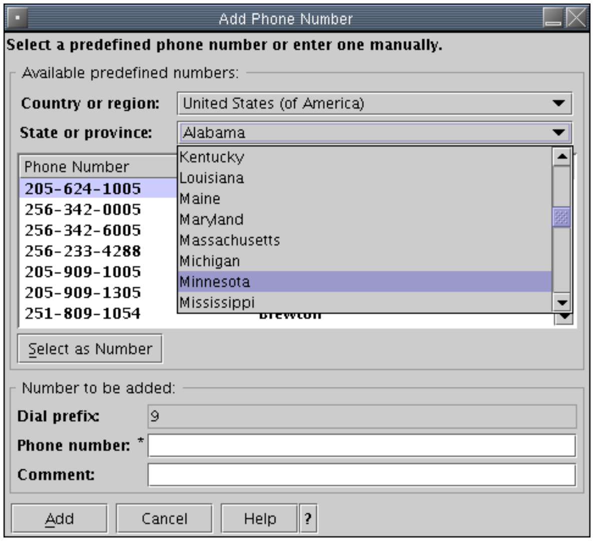
* ☑ Wait for dial tone

* ☑ Enable speaker

Dial prefix: 9

Figure 8-25 Modem configuration

Add the phone number for our country or region. You can add more than one number for backup reason (Figure 8-26, Figure 8-27, and Figure 8-28).



Add Phone Number

Select a predefined phone number or enter one manually.

Available predefined numbers:

Country or region: United States (of America)

State or province: Alabama

Phone Number
205-624-1005
256-342-0005
256-342-6005
256-233-4288
205-909-1005
205-909-1305
251-809-1054

Brewton

Kentucky
Louisiana
Maine
Maryland
Massachusetts
Michigan
Minnesota
Mississippi

Number to be added:

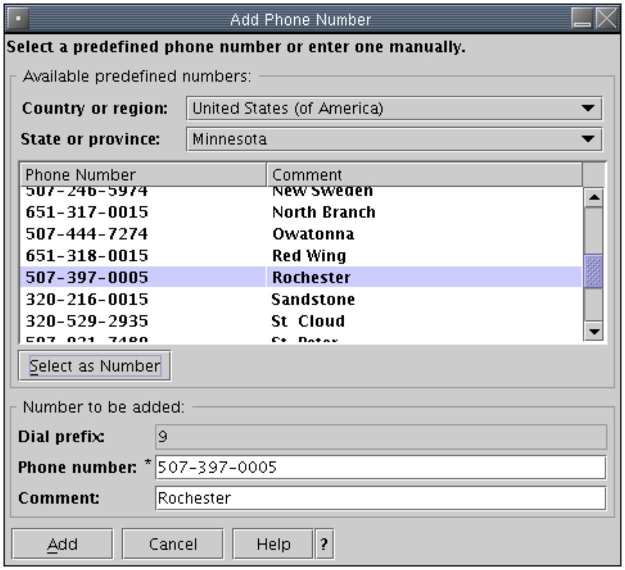
Dial prefix: 9

Phone number: *

Comment:

Figure 8-26 Country or region

After selecting the country, you can add the correct phone number for callhome (Figure 8-27) and other information (Figure 8-28 on page 226).


Add Phone Number

Select a predefined phone number or enter one manually.

Available predefined numbers:

Country or region: United States (of America)

State or province: Minnesota

Phone Number	Comment
507-246-5974	New Sweden
651-317-0015	North Branch
507-444-7274	Owatonna
651-318-0015	Red Wing
507-397-0005	Rochester
320-216-0015	Sandstone
320-529-2935	St Cloud
507-931-7480	St Peter

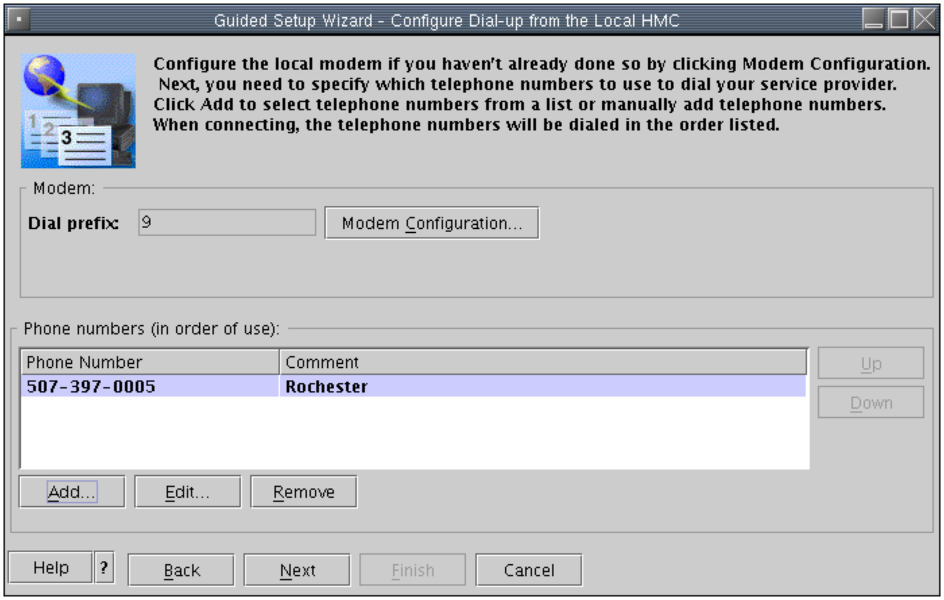
Number to be added:

Dial prefix: 9

Phone number: * 507-397-0005

Comment: Rochester

Figure 8-27 Select phone number for modem


Guided Setup Wizard - Configure Dial-up from the Local HMC

Configure the local modem if you haven't already done so by clicking Modem Configuration. Next, you need to specify which telephone numbers to use to dial your service provider. Click Add to select telephone numbers from a list or manually add telephone numbers. When connecting, the telephone numbers will be dialed in the order listed.

Modem:

Dial prefix: 9

Phone numbers (in order of use):

Phone Number	Comment
507-397-0005	Rochester

Figure 8-28 Dial-up configuration

You are finished with the HMC dialer.

You can also connect to IBM by a VPN channel through your intranet and use an other HMC for dialing or as a gateway to the firewall into the Internet.

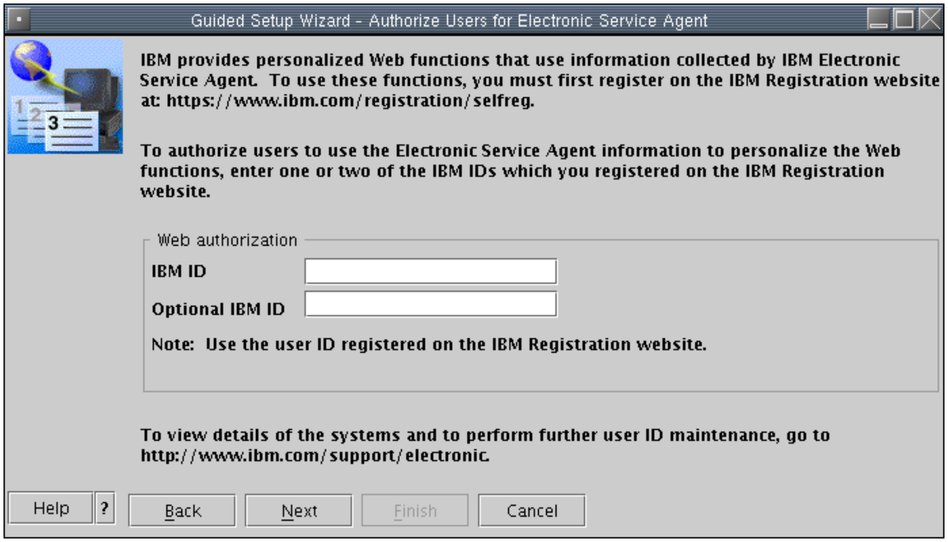
The Guided Setup Wizard - Authorize Users for Electronic Service Agent screen is now displayed (see Figure 2-32). The information collected and sent to IBM by the HMC can be seen on the IBM Electronic Service Agent Web site.

http://www.ibm.com/support/electronic

To access this data on the Web you must have a registered IBM ID and authorized that ID through the HMC. You can register IBM IDs using the Web site.

Enter a valid IBM ID and an optional second IBM ID if required, in the Web authorization panel. The Guided Setup will only allow you to authorize two user IDs to access the data sent by the HMC to IBM. However you can submit as many registrations as you like by clicking Service Applications >>Service Agent >>@server Registration from the HMC desktop (Figure 8-29).

If you do not have valid IBM ID, you can choose to leave this panel blank and manually complete this information later by clicking Service Applications >>Service Agent >> @server Registration from the HMC desktop.



IBM provides personalized Web functions that use information collected by IBM Electronic Service Agent. To use these functions, you must first register on the IBM Registration website at: https://www.ibm.com/registration/selfreg.

To authorize users to use the Electronic Service Agent information to personalize the Web functions, enter one or two of the IBM IDs which you registered on the IBM Registration website.

Web authorization

IBM ID

Optional IBM ID

Note: Use the user ID registered on the IBM Registration website.

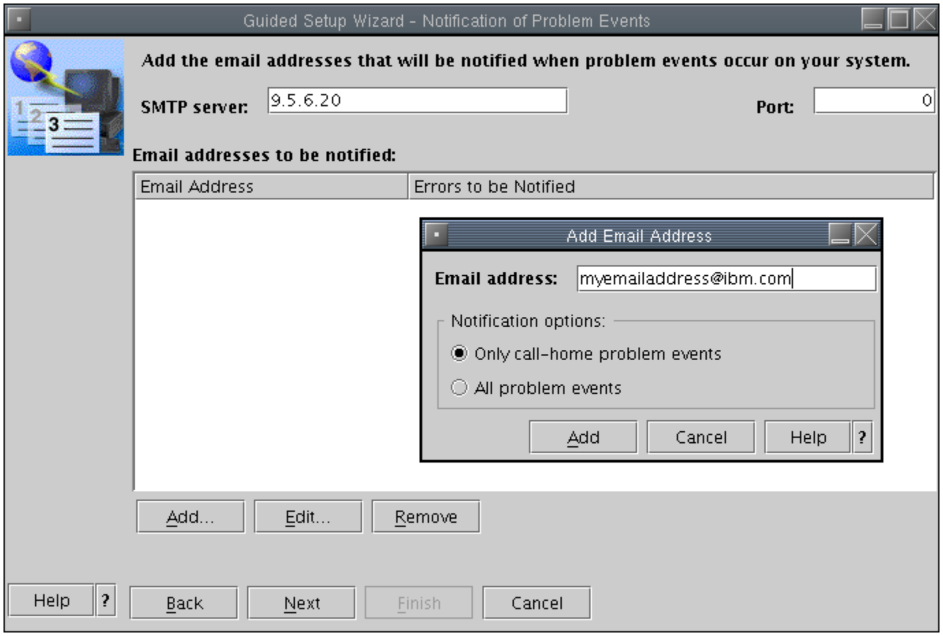
To view details of the systems and to perform further user ID maintenance, go to http://www.ibm.com/support/electronic.

Figure 8-29 Authorized user for ESA

The Guided Setup Wizard - Notification of Problem Events display is shown (Figure 8-30). The HMC can alert your administrators of problems with the HMC or its managed systems using e-mail.

You can choose whether to notify your administrators of only problems reported to IBM (Only call-home problem events) or of all problem events generated.

Enter the IP address and port of your SMTP server. Then click the Add button and enter your administrator's e-mail address and the notification type required. Click Add to accept these values and return to the previous screen. You may enter multiple e-mail addresses by repeating this process.



Guided Setup Wizard - Notification of Problem Events

Add the email addresses that will be notified when problem events occur on your system.

SMTP server: 9.5.6.20

Port: 0

Email addresses to be notified:

Email Address	Errors to be Notified

Add Email Address

Email address: myemailaddress@ibm.com

Notification options:

● Only call-home problem events

○ All problem events

Figure 8-30 The e-mail notification dialog

You can specify how you want the HMC to respond to communications interruptions to its managed system by setting the following values:

- Number of disconnected minutes considered an outage

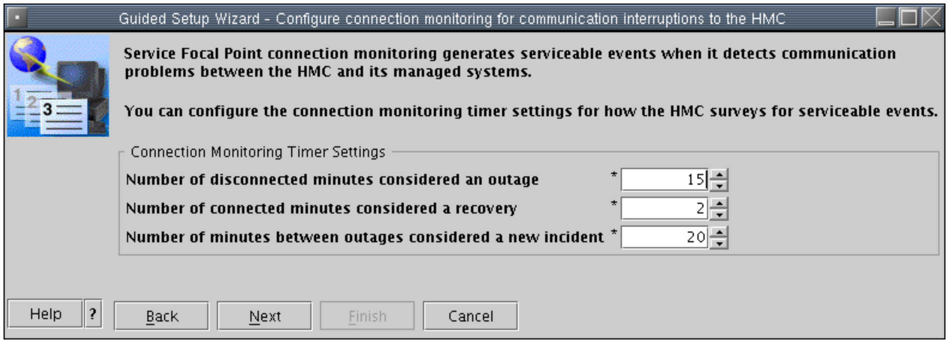
  This is the number of minutes that you want the HMC to wait before reporting a disruption in communications. The recommended time is 15 minutes.

- Number of connected minutes considered a recovery

  This is the number of minutes after communications is restored between the HMC and the managed system that you want the HMC to wait before considering a recovery successful. The recommended time is two minutes.

- Number of minutes between outages considered a new incident

  This is the number of minutes after communication is restored that you want the HMC to wait before considering another outage a new incident. The recommended time is 20 minutes (Figure 8-31).



Guided Setup Wizard - Configure connection monitoring for communication interruptions to the HMC

Service Focal Point connection monitoring generates serviceable events when it detects communication problems between the HMC and its managed systems.

You can configure the connection monitoring timer settings for how the HMC surveys for serviceable events.

Connection Monitoring Timer Settings

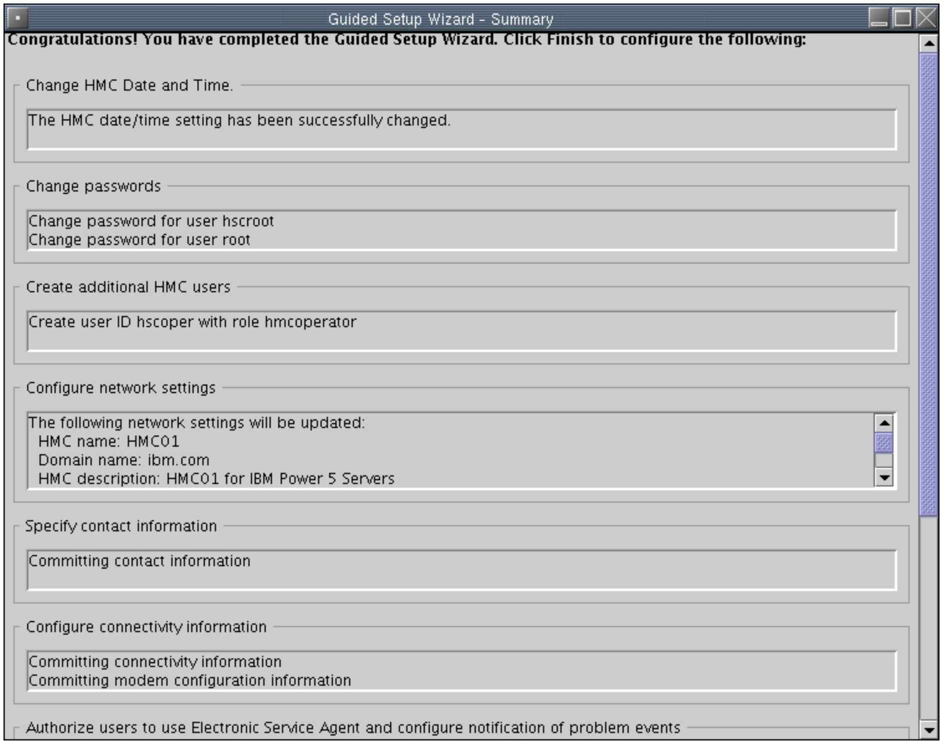
Number of disconnected minutes considered an outage * 15

Number of connected minutes considered a recovery * 2

Number of minutes between outages considered a new incident * 20

Figure 8-31 Communication interruptions

The next screen is the summary screen were you can see all your configuration parameter (Figure 8-32).



Congratulations! You have completed the Guided Setup Wizard. Click Finish to configure the following:

Change HMC Date and Time.

The HMC date/time setting has been successfully changed.

Change passwords

Change password for user hscroot
Change password for user root

Create additional HMC users

Create user ID hscoper with role hmcoperator

Configure network settings

The following network settings will be updated:
HMC name: HMC01
Domain name: ibm.com
HMC description: HMC01 for IBM Power 5 Servers

Specify contact information

Committing contact information

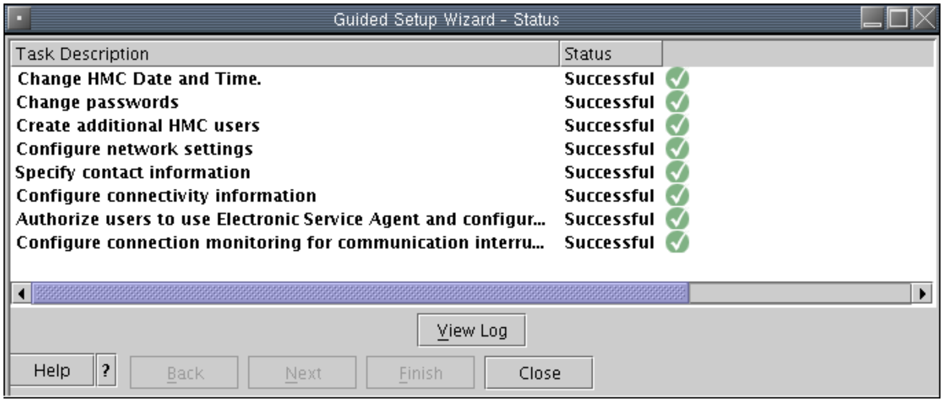
Configure connectivity information

Committing connectivity information
Committing modem configuration information

Authorize users to use Electronic Service Agent and configure notification of problem events

Figure 8-32 Summary screen

Each task which is complete you can see on the status screen (Figure 8-33). It also gives you the option of a status (Figure 2-37).



Task Description	Status
Change HMC Date and Time.	**Successful** ✓
Change passwords	**Successful** ✓
Create additional HMC users	**Successful** ✓
Configure network settings	**Successful** ✓
Specify contact information	**Successful** ✓
Configure connectivity information	**Successful** ✓
Authorize users to use Electronic Service Agent and configur...	**Successful** ✓
Configure connection monitoring for communication interru...	**Successful** ✓

Figure 8-33 Status screen

If any configuration parts were unsuccessful, you can check the status log for configuration problems.

8.6 HMC security and user management

This section discusses security implementation within the HMC environment that includes the following topics:

- Certificate authority
- Server security
- Object manager security
- HMC User management

System Manager Security ensures that the HMC can operate securely in the client-server mode. Managed machines are servers and the managed users are clients. Servers and clients communicate over the Secure Sockets Layer (SSL) protocol, which provides server authentication, data encryption, and data integrity. Each HMC System Manager server has its own private key and a certificate of its public key signed by a Certificate Authority (CA) that is trusted by the System Manager clients. The private key and the server certificate are stored in the server's private key ring file. Each client must have a public key ring file that contains the certificate of the trusted CA.

Define one HMC as a Certificate Authority. You use this HMC to generate keys and certificates for your HMC servers and client systems. The servers are the HMCs you want to manage remotely. A unique key must be generated and installed on each server. You can generate the keys for all your servers in one action on the CA and then copy them to diskette, install them at the servers, and configure the servers for secure operation.

The client systems are the systems from which you want to do remote management. Client systems can be HMCs, AIX, or PC clients. Each client system must have a copy of the CA's public key ring file in its System Manager codebase directory. You can copy the CA public key ring file to the diskette on the CA and copy it from the diskette to each client.

To use the System Manager Security application, you must be a member of the System Administrator role. To ensure security during configuration, users of this application must be logged in to the HMC locally.

Overview and status

The overview and status window displays the following information about the secure system manger server:

- Whether the secure system manager server is configured.
- Whether the private key for this system manager server is installed.
- Whether this system is configured as a Certificate Authority.

Certificate Authority

Define one HMC as a Certificate Authority (CA) to generate keys and certificates for your HMC servers and clients.

A Certificate Authority verifies the identities of the HMC servers to ensure secure

communications between clients and servers. To define a system as a Certificate Authority, you must be logged in as the hscroot user at the machine being defined as the internal Certificate Authority. This procedure defines a system as an internal Certificate Authority for HMC security and creates a public key ring file for the Certificate Authority that you can distribute to all of the clients that access the HMC servers.

A wizard guides you through configuring the Certificate Authority. After you define the internal Certificate Authority, you can use the CA to create the private key files for the HMCs that you want to manage remotely. Each HMC server must have its private key and a certificate of its public key signed by a Certificate Authority that is trusted by the HMC clients. The private key and the server certificate are stored in the server's private key file.There is an option to copy the private key ring files to a diskette so you can install them on your servers.

Note: You cannot perform the following function using a remote client

8.6.1 Server security

This option allows you to install the private key ring file that you have copied to diskette from the HMC server that is acting as the Certificate Authority.Once you have copied the private key file there is another option to configure the HMC as a secure server so that secure, remote clients can be used to remotely manage the HMC.

There is a remote client available for download from the HMC itself. It is called the Web-based System Management remote client and there is a Windows based version and a Linux based version. To run in secure mode a second file needs to be downloaded to the client. This is also available for download from the HMC.

To download the Web-based System Management remote client to your Windows based or Linux based PC, type in the following address from your Web Browser:

#hostname/remote_client.html

where hostname is the name of the HMC you are downloading the Web-based System Management remote client from. You choose whether you want the Windows based version or the Linux based version.

To download the security package so that the client/server, that is, the PC to HMC, connection is secure, type in the following address in your Web Browser:

#hostname/remote_client_security.html

Once again you choose whether you want the Windows based version or the Linux based version.

8.6.2 Object manager security

The HMC Object Manager Security mode can be configured as either Plain Socket or Secure Sockets Layer (SSL). By default, the Plain Sockets mode is used. For SSL mode, the Object Manager reuses the HMC System manager server's private key ring. The server private ring and the Certificate Authority's public key ring must be installed when establishing the SSL connection.

8.6.3 HMC user management

The HMC management option allows you to create and manage HMC user profiles and to configure the HMC. Some of this configuration is done when the setup wizard is initially run to setup the HMC. The options under HMC Management allow you to go in and modify the configuration that was initially set up.

HMC users

This option allows you to perform the following functions for users:

- Creating a user.
- Editing user information.
- Viewing user information.
- Deleting a user.
- Changing passwords.

You must be either the system administrator or the user administrator to perform the functions listed above. Each HMC user can be a member of one to six different roles. Each of these roles allows the user to access different parts of the HMC. The user roles as specified by the HMC are as follows:

- System Administrator
- Advanced Operator
- Service Representative
- Operator
- User Administrator
- Viewer

Each role is described as follows:

- **System Administrator**

  The system administrator acts as the root user, or manager of the HMC system. The system administrator has unrestricted authority to access and modify most of the HMC system.

- **Advanced Operator**

  An advanced operator can perform some partition or system configuration and has access to some user-management functions.

- **Service representative**

  A service representative is the person who installs or repairs the system.

- **Operator**

  An operator is responsible for the daily system operation.

- **User Administrator**

  A user administrator can perform user management tasks, but cannot perform any other HMC functions.

- **Viewer**

  A viewer can view HMC information, but cannot change any configuration information.

A user ID also needs to be created for the software service representative so they have access to perform fixes on the HMC code. This user ID is hscpe and it must not be assigned to any other user. This user name is reserved for your support representative and is considered a hidden role.

8.7 Inventory Scout services

The Inventory Scout is a tool that surveys managed systems for hardware and software information. Inventory Scout provides an automatic configuration mechanism and eliminates the need for you to manually reconfigure Inventory Scout Services. Depending on the levels of your HMC and partition software, you might be required to manually configure partitions that you create in order to perform Inventory Scout tasks. The Inventory Scout collects the Vital Product Data (VPD) from the hardware resources in the managed system or systems that the HMC is managing. For Inventory Scout to collect all the information accurately all the managed system partitions must be active. The Inventory Scout collects such information as a resource type and serial number, its part number, its operational status and other VPD depending on the resource type.

This VPD is sent weekly to a database at IBM by a scheduled job. The initial collection of data will send all collected VPD but any subsequent transmission will only send what has changed since the last transmission to IBM. The information sent to IBM is a valuable aid for IBM Remote Technical Support personnel. When solving problems on the managed systems as they will have an accurate profile of what the resources are on the system without having to connect to the system.

There are three options available under Inventory Scout services:

- Inventory scout profile configuration.
- Collect VPD information.
- Restart inventory scout daemon.

Inventory Scout profile configuration

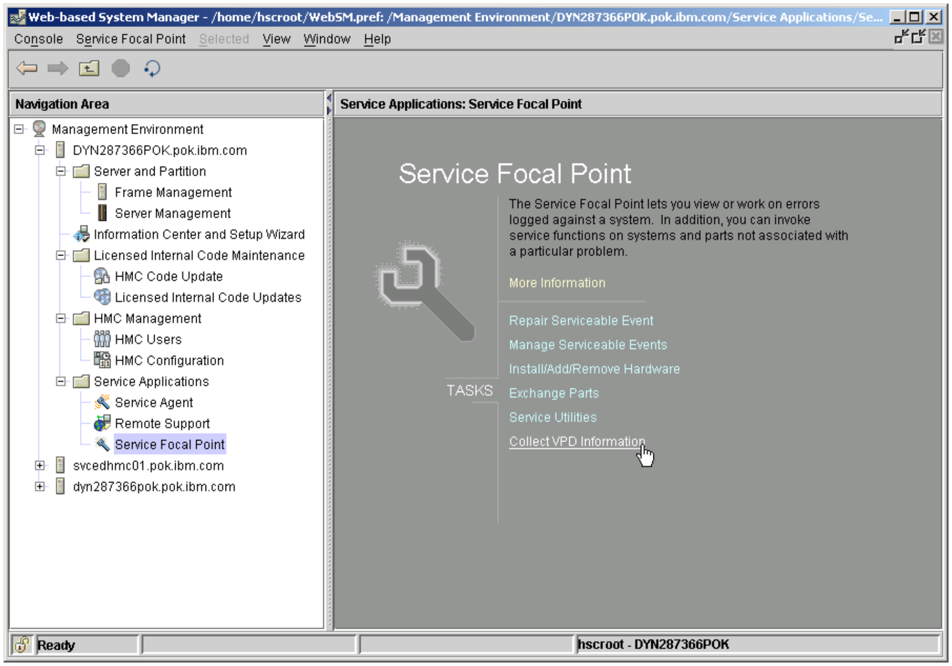
Running this option runs a wizard which guides you through setting up the Inventory Scout profile. This should only be needed to run if the initial setup wizard for the HMC was not run or if a new AIX partition has been created since the initial setup wizard was run on the HMC.

Collect VPD information - This collects the VPD to diskette if required. Restart inventory scout daemon - Option to restart the inventory scout daemon. For this the invscout must run on the partition. check this with:

```
#ps -ef | grep envscoutd
```

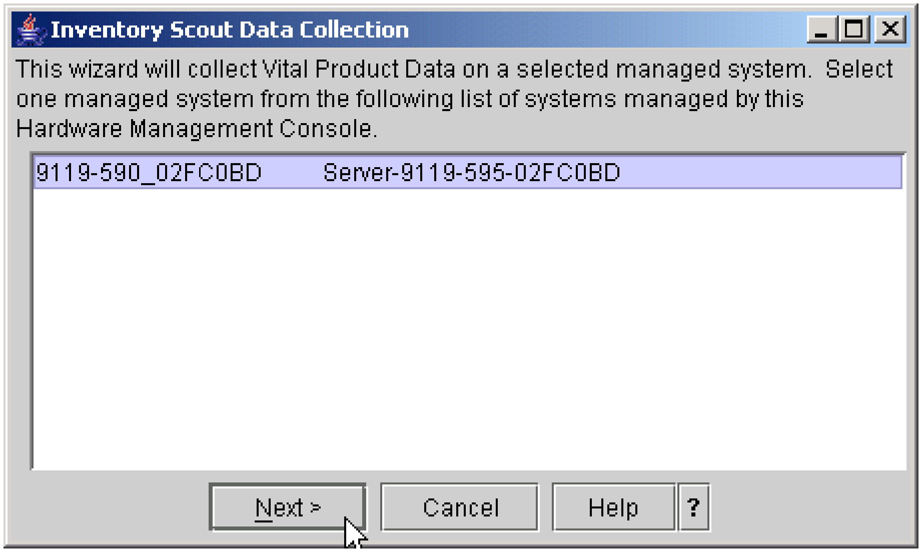
and restart with: Than you can also check the version auf the inventory scout.

```
#invscout
```



The Service Focal Point lets you view or work on errors logged against a system. In addition, you can invoke service functions on systems and parts not associated with a particular problem.

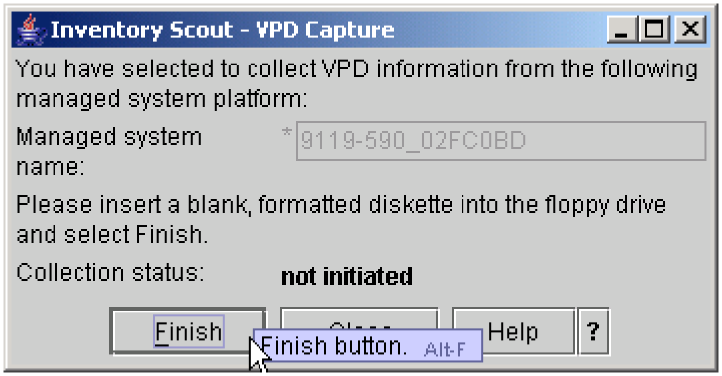
Figure 8-34 Inventory Scout



This wizard will collect Vital Product Data on a selected managed system. Select one managed system from the following list of systems managed by this Hardware Management Console.

9119-590_02FC0BD Server-9119-595-02FC0BD

Figure 8-35 Select server to get VPD data



Inventory Scout - VPD Capture

You have selected to collect VPD information from the following managed system platform:

Managed system name: * 9119-590_02FC0BD

Please insert a blank, formatted diskette into the floppy drive and select Finish.

Collection status: **not initiated**

Figure 8-36 Store data

8.8 Service Agent and Service Focal Point

Remote support enables connectivity to IBM from the HMC. The 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 via 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 options available for remote support are as follows;

- 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 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 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.

8.8.1 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.

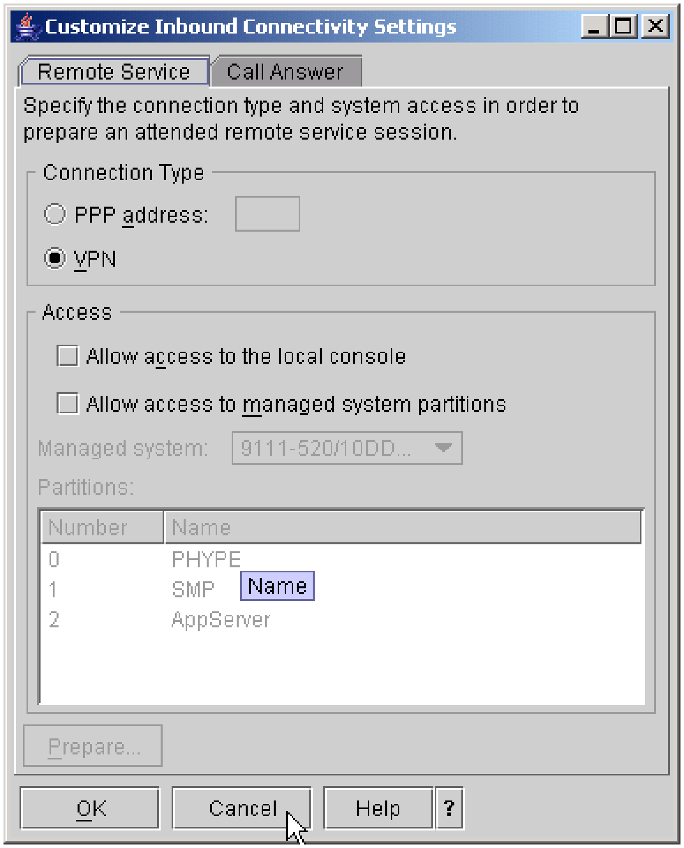
For the client we can define E-mail notification to the E-mail account.

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.



Customize Inbound Connectivity Settings

Remote Service | Call Answer

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 8-37 PPP or VPN connection

As you can see, this is a VPN connection to IBM

8.8.2 Service Focal Point

The Service Focal Point 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 a Ethernet connection HMC to each LPAR.

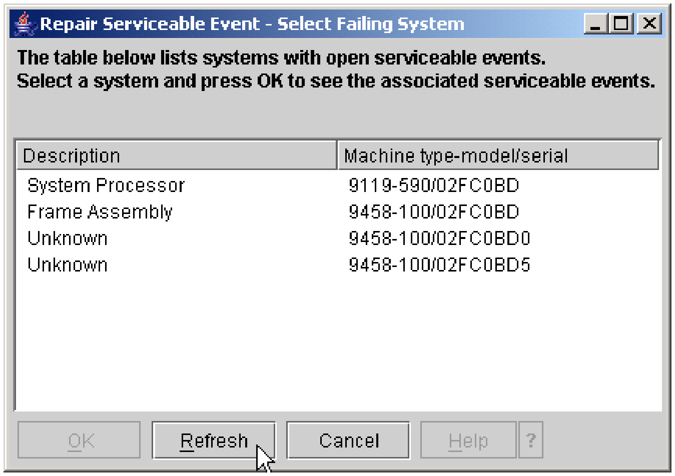
The options available under Service Focal Point are as follows:

- Repair serviceable event.
- Manage serviceable events.
- Install/add/remove hardware.
- Replace parts.
- Service utilities.

Repair serviceable event

This option allows the user or the service representative to view a serviceable event and then initiate a repair against that service event. The following is 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 allows a user to view the events and perform problem determination.

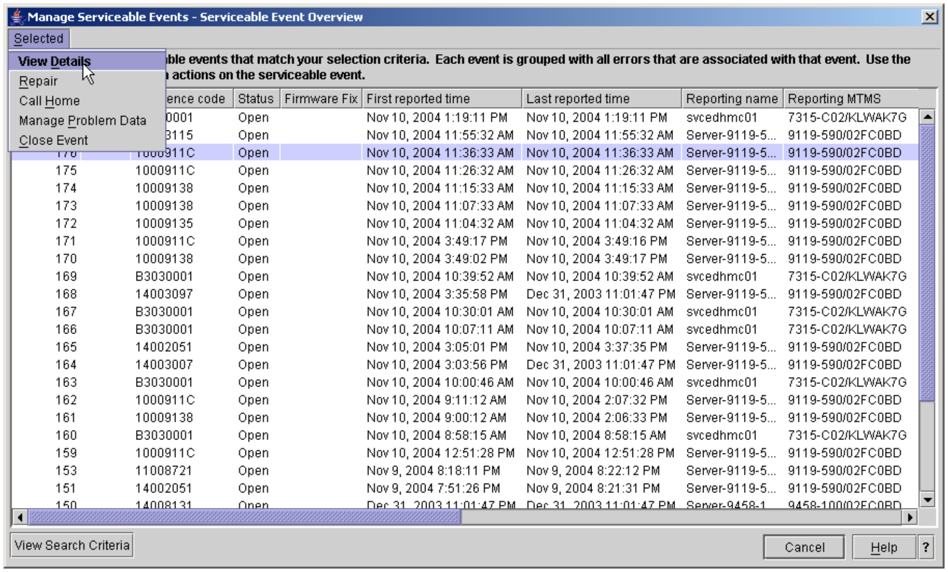


The table below lists systems with open serviceable events.
Select a system and press OK to see the associated serviceable events.

Description	Machine type-model/serial
System Processor	9119-590/02FC0BD
Frame Assembly	9458-100/02FC0BD
Unknown	9458-100/02FC0BD0
Unknown	9458-100/02FC0BD5

Figure 8-38 Open serviceable events

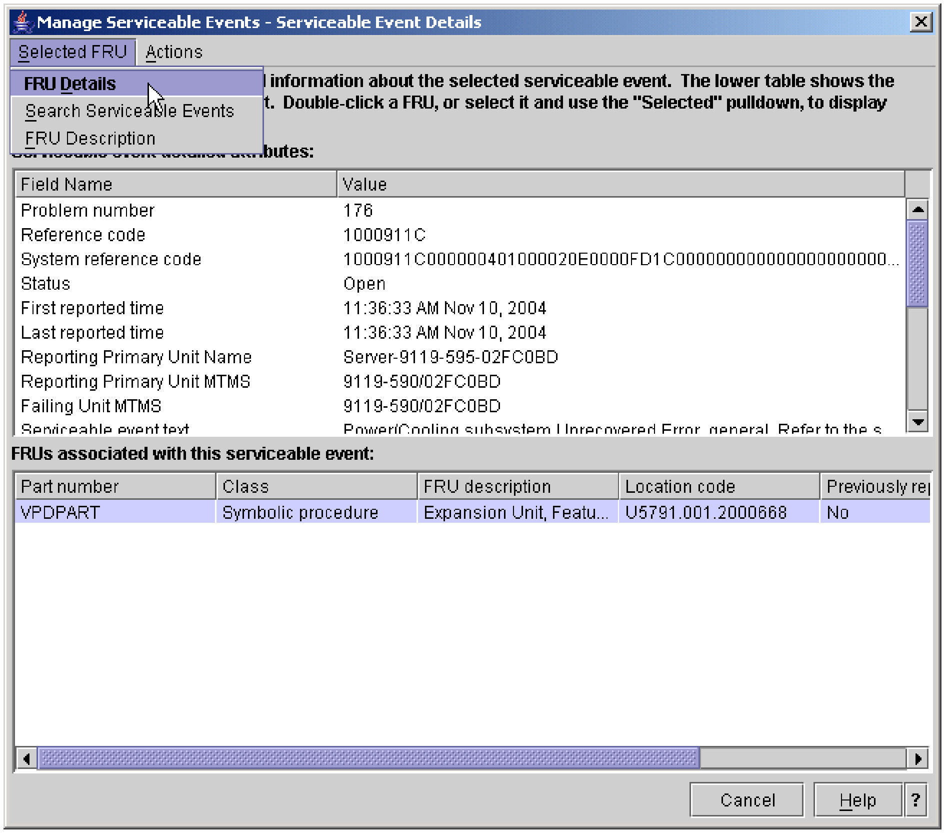
On this page we can select the service events to see more data.



ble events that match your selection criteria. Each event is grouped with all errors that are associated with that event. Use the
actions on the serviceable event.

	ence code	Status	Firmware Fix	First reported time	Last reported time	Reporting name	Reporting MTMS
	0001	Open		Nov 10, 2004 1:19:11 PM	Nov 10, 2004 1:19:11 PM	svcedhmc01	7315-C02/KLWAK7G
	3115	Open		Nov 10, 2004 11:55:32 AM	Nov 10, 2004 11:55:32 AM	Server-9119-5...	9119-590/02FC0BD
176	1000911C	Open		Nov 10, 2004 11:36:33 AM	Nov 10, 2004 11:36:33 AM	Server-9119-5...	9119-590/02FC0BD
175	1000911C	Open		Nov 10, 2004 11:26:32 AM	Nov 10, 2004 11:26:32 AM	Server-9119-5...	9119-590/02FC0BD
174	10009138	Open		Nov 10, 2004 11:15:33 AM	Nov 10, 2004 11:15:33 AM	Server-9119-5...	9119-590/02FC0BD
173	10009138	Open		Nov 10, 2004 11:07:33 AM	Nov 10, 2004 11:07:33 AM	Server-9119-5...	9119-590/02FC0BD
172	10009135	Open		Nov 10, 2004 11:04:32 AM	Nov 10, 2004 11:04:32 AM	Server-9119-5...	9119-590/02FC0BD
171	1000911C	Open		Nov 10, 2004 3:49:17 PM	Nov 10, 2004 3:49:16 PM	Server-9119-5...	9119-590/02FC0BD
170	10009138	Open		Nov 10, 2004 3:49:02 PM	Nov 10, 2004 3:49:17 PM	Server-9119-5...	9119-590/02FC0BD
169	B3030001	Open		Nov 10, 2004 10:39:52 AM	Nov 10, 2004 10:39:52 AM	svcedhmc01	7315-C02/KLWAK7G
168	14003097	Open		Nov 10, 2004 3:35:58 PM	Dec 31, 2003 11:01:47 PM	Server-9119-5...	9119-590/02FC0BD
167	B3030001	Open		Nov 10, 2004 10:30:01 AM	Nov 10, 2004 10:30:01 AM	svcedhmc01	7315-C02/KLWAK7G
166	B3030001	Open		Nov 10, 2004 10:07:11 AM	Nov 10, 2004 10:07:11 AM	svcedhmc01	7315-C02/KLWAK7G
165	14002051	Open		Nov 10, 2004 3:05:01 PM	Nov 10, 2004 3:37:35 PM	Server-9119-5...	9119-590/02FC0BD
164	14003007	Open		Nov 10, 2004 3:03:56 PM	Dec 31, 2003 11:01:47 PM	Server-9119-5...	9119-590/02FC0BD
163	B3030001	Open		Nov 10, 2004 10:00:46 AM	Nov 10, 2004 10:00:46 AM	svcedhmc01	7315-C02/KLWAK7G
162	1000911C	Open		Nov 10, 2004 9:11:12 AM	Nov 10, 2004 2:07:32 PM	Server-9119-5...	9119-590/02FC0BD
161	10009138	Open		Nov 10, 2004 9:00:12 AM	Nov 10, 2004 2:06:33 PM	Server-9119-5...	9119-590/02FC0BD
160	B3030001	Open		Nov 10, 2004 8:58:15 AM	Nov 10, 2004 8:58:15 AM	svcedhmc01	7315-C02/KLWAK7G
159	1000911C	Open		Nov 10, 2004 12:51:28 PM	Nov 10, 2004 12:51:28 PM	Server-9119-5...	9119-590/02FC0BD
153	11008721	Open		Nov 9, 2004 8:18:11 PM	Nov 9, 2004 8:22:12 PM	Server-9119-5...	9119-590/02FC0BD
151	14002051	Open		Nov 9, 2004 7:51:26 PM	Nov 9, 2004 8:21:31 PM	Server-9119-5...	9119-590/02FC0BD
150	14008131	Open		Dec 31, 2003 11:01:47 PM	Dec 31, 2003 11:01:47 PM	Server-9458-1...	9458-100/02FC0BD

Figure 8-39 Manage service events


Manage Serviceable Events - Serviceable Event Details
Selected FRU
Actions
FRU Details
Search Serviceable Events
FRU Description
Cancel
Help
?

information about the selected serviceable event. The lower table shows the t. Double-click a FRU, or select it and use the "Selected" pulldown, to display

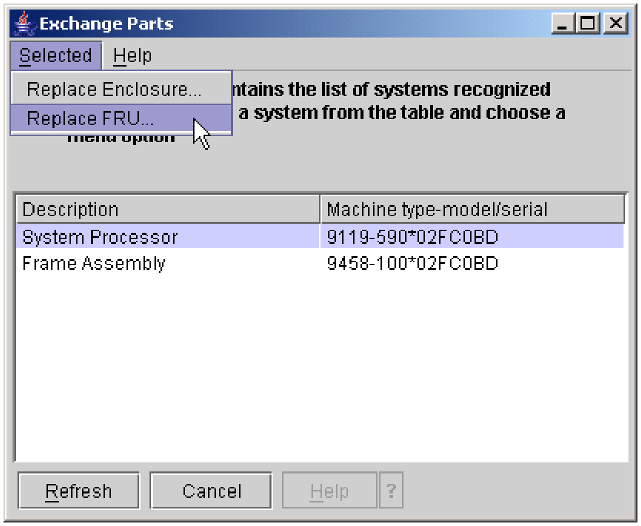
butes:

Field Name	Value
Problem number	176
Reference code	1000911C
System reference code	1000911C000000401000020E0000FD1C00000000000000000000...
Status	Open
First reported time	11:36:33 AM Nov 10, 2004
Last reported time	11:36:33 AM Nov 10, 2004
Reporting Primary Unit Name	Server-9119-595-02FC0BD
Reporting Primary Unit MTMS	9119-590/02FC0BD
Failing Unit MTMS	9119-590/02FC0BD
Serviceable event text	Power/Cooling subsystem Unrecovered Error, general. Refer to the s

FRUs associated with this serviceable event:

Part number	Class	FRU description	Location code	Previously re
VPDPART	Symbolic procedure	Expansion Unit, Featu...	U5791.001.2000668	No

Figure 8-40 Detail view of a service event



ntains the list of systems recognized
a system from the table and choose a
menu option

Description	Machine type-model/serial
System Processor	9119-590*02FC0BD
Frame Assembly	9458-100*02FC0BD

Figure 8-41 Exchange parts

8.9 HMC service utilities and tasks

The Advanced System Management Interface (ASMI) is a service that is provided to allow access to the system to perform actions including powering on the system, changing the boot configuration, and other tasks. ASMI can be launched from a Web browser or from the HMC. Section 7.3, “Advanced System Management Interface (ASMI)” on page 175 also covers ways to access the ASMI menus.

8.9.1 HMC boot up fails with fsck required

There are certain circumstances where the HMC can fail and leave the file subsystem in an unknown state. Power failure is the most common form of incident, caused by inadvertently pressing the power button during a disk access, pulling the power cable, or power supply failure.

The HMC has the ability to recover itself in most incidents. There may be a time when the HMC stops the boot process with the following error message displayed:

```
fsck failed please repair manually. login with root
```

If you have completed the Guided Setup you will have changed the default root password. You should use this new password.

Once you have logged in with root password you will be presented with a command prompt. At the prompt Enter the following command.

```
#fsck
```

You will see the **fsck** command process running in line mode. You will be asked if the OS should repair certain component. You should answer yes to all prompts unless you are instructed otherwise. Once the **fsck** command completes, the HMC should automatically return to the GUI signon screen. If you are not returned to the GUI, you should Enter the following command.

```
#reboot
```

The HMC GUI signon screen will appear.

8.9.2 Determining HMC serial number

For some HMC or service processor troubleshooting situations a Product Engineer (PE) will have to sign into the HMC. The PE password changes daily and is not available for normal client use. If the PE determines a local service engineer can sign on to the HMC the PE may request the HMC serial number.

To find the HMC serial number, open a restricted shell window and run the following command,

```
#lshmc -v
```

A

Facts and features reference

The following section contains an extract of the facts and features reference, which is available from the following Web site:

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

Use this section as a general reference and be aware that the information located on the Web may include updates that this publication does not reflect.

Table A-1 Facts and Features for p5-590 and p5-595

Product line	**p5-590**	**p5-595**
Machine type-model	9119-590	9119-595
System packaging	24” system frame (+ expansion frame)	24” system frame (+ expansion frame)
Microprocessor Type	64-bit POWER5	64-bit POWER5
# of processors/system	8 to 32	16 to 64
Clock rates available	1.65 GHz	1.65 GHz, 1.9 GHz
System memory (standard/maximum)	8 GB-1024 GB [a,h]	8 GB-2048 GB [a,i]
Data/instruction (L1) cache	32 KB/64 KB [b]	32 KB/64 KB [b]
Level 2 (L2) cache	7.6 MB [d]	7.6 MB [d]
Level 3 (L3) cache	144 MB [d]	144 MB [d]
Reliability, availability, serviceability		
Chipkill memory	X	X
Service processor	X [f]	X [f]
Hot-swappable disks (internal and external)	X	X
Dynamic Processor Deallocation	X	X
Dynamic deallocation: PCI bus slots	X	X
Hot-plug slots	X	X
Blind-swap slots	X	X
Redundant hot-plug power	X	X
Redundant hot-plug cooling	X	X
NEBS3	-	-
Capacity		
Capacity on Demand (CoD) features	P, M, R, B (SOD), OOP, OOM	P, M, R, B (SOD), OOP, OOM
Maximum logical partitions/micro-partitions	254	254
Advanced POWER Virtualization	X	X
Maximum available PCI slots	160 PCI (64-bit) [c]	240 PCI (64-bit) [c]
Maximum PCI bus speed	133 MHz	133 MHz
Disk | media bays	128 [c] | -	192 [c] | -
Minimum | maximum internal disk storage [c]	72.8 GB | 9.3 TB	72.8 GB | 14.0 TB
Required | optional I/O drawers	1 | 7	1 | 11
Storage interfaces (maximum [c,g])		
FC 6204 Ultra SCSI Differential - PCI	32	32
FC 6203 2-Channel Ultra3 SCSI - PCI	-	-
FC 5710 or FC 5712 2-channel Ultra320 SCSI - PCI-X	62	62
FC 5703 or FC 5711 2-channel Ultra320 RAID - PCI-X	62	62
FC 6230 SSA Advanced SerialRAID Plus	-	-
FC 6239 or FC 5716 2 Gigabit Fibre Channel - PCI-X	160	192
Communications and Connectivity (maximum [c,g])		
FC 2738 or FC 2962 - 2/4-port Multiprotocol PCI	16	16
FC 2943 - 8-port Asynchronous - EIA-232 / RS-422	32	32
FC 2944 - 128-port Asynchronous Controller	32	32
FC 4959 - Token-Ring 4/16 Mbps	20	20
FC 4961 - 4-port 10/100 Mbps Ethernet	-	-
FC 4962 - 10/100 Mbps Ethernet PCI II	160	192
FC 5701 - 10/100/1000 Mbps Base-TX Ethernet PCI-X	160	192
FC 5706 - 2-port 10/100/1000 Mbps Base-TX Ethernet	160	192
FC 5700 - Gigabit Ethernet-SX PCI-X (Fibre)	160	192
FC 5707 - 2-port Gigabit Ethernet-SX PCI-X (Fibre)	160	192
FC 5718 - 10 Gigabit Ethernet - PCI-X	24	24
FC 8398 - SP Switch2 System Attachment	-	-
pSeries High Performance Switch Network Interface	SOD	SOD
FC 2751 - ESCON Channel PCI [e]	-	-
FC 2732 - HiPPI [e]	-	-
FC 6312 - Quad Digital Trunk Telephony PCI [e]	-	-
FC 4960 - e-business Cryptographic Accelerator	-	-
FC 4963 - PCI Cryptographic Coprocessor (FIPS-4)	-	-
Display adapter (maximum [c,g])	GXT135P (16)	GXT135P (16)

Footnotes

- X = Standard; Supported
- O = Optionally Available; Supported
- - = Not Applicable
- P = Processor Capacity Upgrade on Demand option
- B = Capacity BackUp offering
- M = Memory Capacity Upgrade on Demand option
- OOP = On/Off Capacity on Demand for Processors option
- OOM = On/Off Capacity on Demand for Memory option
- R = Reserve Capacity in Demand option
- SOD = Statement of General Direction announced

a Shared memory

b Per processor

c Using maximum number of I/O drawers

d Per processor card, processor book or Multichip Module

e Requires additional software; check on availability

f Statement of Direction for redundant service processor - 1H 2005

g More details on interface and adapter dependencies and their effect on these maximums can be found in the IBM Sales Manual or the pSeries PCI Adapter Placement manual available at:
http://publib16.boulder.ibm.com/pseries/en_US/infocenter/base/HW_pci_adp_pl.htm

h Figures for DDR1 266 MHz memory; For DDR2 533 MHz memory, range is 8 GB to 128 GB

i Supported on DDR1 memory only

Table A-2 System unit details

System unit details	p5-590	p5-595
Standard internal disk bays	16[1]	16[1]
Optional I/O drawer disk bays	112	112
Available media bays	O	O
- Standard size	-	-
- Slimline size	-	-
Standard diskette drive	O	O
Standard DVD-ROM	O	O
Integrated RS232 serial ports	-	-
Integrated USB ports	-	-
Keyboard/Mouse Port	-	-
Parallel port	-	-
HMC ports	2	2
Integrated 10/100 Ethernet port	-	-
Integrated 10/100/1000 Ethernet port	-	-
Integrated SCSI port / controller	2[1]	2[1]
- Max SCSI speed (Mbps)	160	160
PCI slots	20[1]	20[1]
- Long 64-bit 5v 33 MHz	6	6
- Long 64-bit 3.3v 50 / 66 MHz	14	14
PCI-X slots[2]	20[1]	20[1]
- Long 64-bit 3.3v 133 MHz	20	20
- Short 64-bit 3.3v 133 MHz	-	-
- Short 32-bit 3.3v 66 MHz	-	-
RJ-4x connector	-	-
Rack light indicator	-	-
LED diagnostics	X	X

X= Available; - = Not Available; O= Optional

[1] Assuming single required I/O drawer; either PCI or PCI-X slot drawer may be installed.

[2] Assuming optional I/O drawers are not installed

Table A-3 Server I/O attachment

Server I/O attachment	Max. Drawer per System	Slots per Drawer	Max. Slots per System	Disk Bays per Drawer	Max. Disk Bays per System	Max. I/O Drawer Disk Capacity	Max. Disk Capacity per System
p5-590[1]	8		160		128		9.3 TB
FC 5791 (internal drawer)		20 PCI-X		16		9.3 TB	
FC 5794 (internal drawer)		20 PCI-X		8		7.0 TB	
7040-61D drawer[2]	3[1]	20 PCI or 20 PCI-X		16		9.3 TB	
p5-595[1]	12		240		192		14.0 TB
FC 5791 (internal drawer)		20 PCI-X		16		14.0 TB	
FC 5794 (internal drawer)		20 PCI-X		8		7.0 TB	
7040-61D drawer[2]		20 PCI or 20 PCI-X		16		14.0 TB	

[1] At least one drawer is required
[2] Ultra320 SCSI adapter provides access to external Ultra320 disk drives

Table A-4 Peak bandwidth

Peak bandwidth	**p5-590**	**p5-595**
Memory to processor (GB/second)	399.7*	799.5*
L2 to L3 cache (GB/second)	422.4	972.8
RIO-2 I/O subsystem (GB/second)	48	72

* Using 533 MHz DDR2 memory

Table A-5 Standard warranty in United States, other countries may vary

Standard warranty	**p5-590**	**p5-595**
24 x 7 with same-day service objective / Customer replaceable units (CRU)	X	X

Table A-6 Physical planning characteristics

Server	**p5-590**		**p5-595**	
	Min.	Max.	Min.	Max.
Number of processors	8-way	32-way	16-way	64-way
Packaging	Frame(s) w/drawers	Frame(s) w/drawers	Frame(s) w/drawers	Frame(s) w/drawers
KVA	-	22.7	-	22.7
Watts	-	22710	-	22710
BTU/hour	-	77500	-	77500
Noise (bels)	7.6	8.3	7.6	8.3
Voltage	200-240, 380-415, 480 3-phase		200-240, 380-415, 480 3-phase	
Power supply	N+1 Standard; IBB optional		N+1 Standard; IBB optional	
Height inches	42U - 79.7		42U - 79.7	
millimeters	2025		2025	
Width inches	62.0		62.0	
millimeters	1575		1575	
Depth inches	52.2 - 62.2		52.2 - 62.2	
millimeters	1326 - 1681		1326 - 1681	
Operating Temperature (°C)	10 - 32		10 - 32	10 - 28
Operating Humidity	8% - 80%		8% - 80%	
Max. Altitude feet	10000		10000	7000
meters	3048		3048	2135
Weight pounds	2735	4956	2735	5420
kilograms	1241	2248	1241	2458

Table A-7 Racks

Racks		7014-T00	7014-T42	7040-61R
		36U	42U	42U
Height	inches	71.0 - 75.5	79.3	79.72
	millimeters	1804 - 1926	2015	2025
Width	inches	24.5 - 25.4	24.5 - 25.4	30.91
	millimeters	623 - 644	623 - 644	785
Depth	inches	41.0 - 45.2	41.0 - 45.2	52.82 - 58.83
	millimeters	1042 - 1098	1042 - 1098	1342 - 1494
Weight	pounds	535	575	-
	kilograms	244	261	-

Table A-8 I/O device options list

I/O device options	p5-590 / p5-595
Disk drives and subsystems	
FC 7204 External disk drive	X
FC 7133 Serial disk system	X
FC 2104 Expandable storage plus U320	X
FC 2105 Enterprise storage server	X
FC 5198 NAS gateway 500	X
DS4000 series disk systems	X
DS6000 series disk systems	X
DS8000 series disk systems	X
Fibre channel directors, switches, hubs	
FC 2031 McDATA fibre channel switch	X
FC 2032 McDATA fibre channel director	X
FC 2042 INRANGE FC/9000 fibre channel	X
FC 2062 Cisco Fabric Switch / Director	X
FC 3534 SAN switch	X
FC 2109 SAN switch	X
Optical drives and libraries	
4.7GB Auto-docking DVD-RAM drive	-
7210-025/030 DVD-RAM drive	X
FC 2634 16X/48X IDE DVD-ROM drive	-
Tape drives and libraries	
FC 3494 Tape Library	X
FC 3580 Tape Drive	X
FC 3581 Tape Autoloader	X
FC 3582 Tape Library	X
FC 3583 Tape Library	-
FC 3584 Tape Library	X
FC 3590 Tape Drive	X
FC 3592 Tape Drive	X
FC 7205 External DLT Drive	X
FC 7206 4mm Tape Drive	X
FC 7207 4 GB 1/4-inch External Cartridge Tape	X
FC 7208 External 8mm Tape Drive	X
FC 7212 SLR60/SLR100 Tape Drives	X
FC 7332 4mm DDS-3 Tape Autoloader	X
FC 6158 20/40 GB 4 mm Tape Drive	-
FC 6134 60/150 GB 8 mm Tape Drive (internal)	-
FC 6120 80/160 GB VXA Tape Drive (internal)	-
FC 6256 36/72 GB 4mm DAT72 Tape Drive (internal)	-

B

PCI adapter placement guide

PCI adapters connected to the model 590 or 595 pSeries system units are placed in expansion units. The following information provides direction on what adapters can be placed in the 5791, 5794, and 61D expansion units and where adapters should be placed for optimum performance. This information will be removed in later editions of this guide, as the Information Center is frequently updated.

Note:

- Model 5791 and 5794 expansion units are orderable.
- Model 61D expansion units can be migrated if they contain the PCI-X planar (FC 6571). Units with the non-PCI-X planar (FC 6563) cannot be migrated.

Expansion unit back view PCI

Figure B-1 on page 254 shows 61D expansion unit back view with numbered slots.

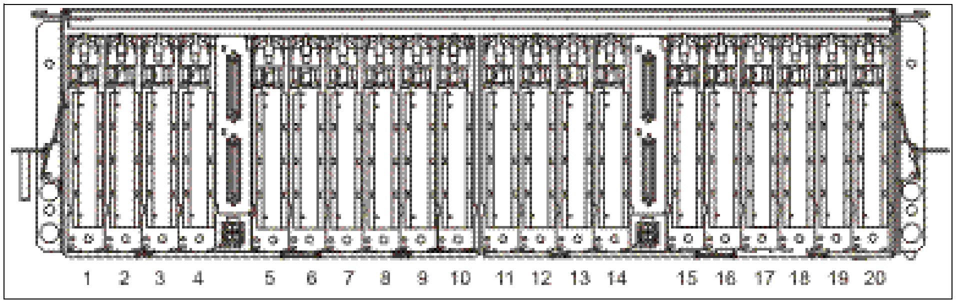


Figure B-1 7040-61D expansion unit back view with numbered slots

PCI slot description

The following tables show the slot properties and PHB connections:

Table B-1 Model 61D expansion unit slot location description (PHB 1 and 2)

	PHB0				PHB2			
Planar1	1	2	3	4	5	6	7	Integrated SCSI U160
Planar2	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	

Table B-2 Model 61D expansion unit slot location description (PHB 3)

	PHB3			
Planar1	8	9	10	Integrated SCSI U160
Planar2	18	19	20	
	Long	Long	Long	66 MHz
	64-bit 3.3V, 133 MHz	64-bit 3.3V, 133 MHz	64-bit 3.3V, 133 MHz	

Slots 1 through 20 are compatible with PCI or PCI-X adapters
Short adapters can go in short or long slots.
All slots support Enhanced Error Handling (EEH)

Note:

The Uffff.ccc.sssssss.Pn.Cm..... represents the Hardware Management Console (HMC) location code, which provides information as to the identify of the enclosure, backplane, PCI adapter(s), and connector. The ffff.ccc.sssssss in the location code represents the following:

- ffff = Feature Code of the Enclosure (drawer or node)
- ccc = the Sequence Number of the Enclosure
- sssssss = the Serial Number of the Enclosure.

Recommended system unit slot placement and maximums

Table B-3 p5-590 and p5-595 PCI adapter placement table

Feature code	Expansion unit slot priority	Expansion unit maximum	System maximum
		These maximums are for connectivity	
FC 5719**	1, 11, 5, 15, 8, 18, 2, 12, 6, 16, 9, 19, 3, 13, 4, 14, 7, 17, 10, 20	4	24 for Model 595 16 for Model 590
FC 5718**	1, 11, 5, 15, 8, 18, 2, 12, 6, 16, 9, 19, 3, 13, 4, 14, 7, 17, 10, 20	4	24 for Model 595 16 for Model 590
FC 5707*	1, 11, 6, 16, 9, 19, 2, 12, 7, 17, 10, 20, 3, 13, 4, 14	16	192 for Model 595 128 for Model 590
FC 5706*	1, 11, 5, 15, 8, 18, 2, 12, 6, 16, 9, 19, 3, 13, 4, 14, 7, 17, 10, 20	20	192 for Model 595 160 for Model 590
FC 5701*	1, 11, 5, 15, 8, 18, 2, 12, 6, 16, 9, 19, 3, 13, 4, 14, 7, 17, 10, 20	20	192 for Model 595 160 for Model 590
FC 5700*	1, 11, 5, 15, 8, 18, 2, 12, 6, 16, 9, 19, 3, 13, 4, 14, 7, 17, 10, 20	20	192 for Model 595 160 for Model 590
FC 5714*	1, 11, 5, 15, 8, 18, 2, 12, 6, 16, 9, 19, 3, 13, 4, 14, 7, 17, 10, 20	20	48

Feature code	Expansion unit slot priority	Expansion unit maximum	System maximum
FC 5713*	1, 11, 5, 15, 8, 18, 2, 12, 6, 16, 9, 19, 3, 13, 4, 14, 7, 17, 10, 20	20	48
FC 5716*	1, 11, 5, 15, 8, 18, 2, 12, 6, 16, 9, 19, 3, 13, 4, 14, 7, 17, 10, 20	20	192 for Model 595 160 for Model 590
FC 5710*	1, 11, 5, 15, 8, 18, 2, 12, 6, 16, 9, 19, 3, 13, 4, 14, 7, 17, 10, 20	20	62
FC 5711*	1, 11, 5, 15, 8, 18, 2, 12, 6, 16, 9, 19, 3, 13, 4, 14, 7, 17, 10, 20	20	62
FC 6230*	1, 11, 5, 15, 8, 18, 2, 12, 6, 16, 9, 19, 3, 13, 4, 14, 7, 17, 10, 20	20	62
FC 6231	128 MB DRAM Option Card for FC 6230		
FC 6235	32 MB Fast-Write Cache Option Card for FC 6230		
FC 2738	10, 20, 9, 19, 8, 18, 7, 17, 6, 16, 5, 15, 4, 14, 3, 13, 2, 12, 1, 11	4	16
FC 2849	10, 20, 9, 19, 8, 18, 7, 17, 6, 16, 5, 15, 4, 14, 3, 13, 2, 12, 1, 11	4	16
FC 2943	10, 20, 9, 19, 8, 18, 7, 17, 6, 16, 5, 15, 4, 14, 3, 13, 2, 12, 1, 11	20	32
FC 2944	10, 20, 9, 19, 8, 18, 7, 17, 6, 16, 5, 15, 4, 14, 3, 13, 2, 12, 1, 11	20	32
FC 2947	10, 20, 9, 19, 8, 18, 7, 17, 6, 16, 5, 15, 4, 14, 3, 13, 2, 12, 1, 11	16	16

Feature code	Expansion unit slot priority	Expansion unit maximum	System maximum
FC 2962	10, 20, 9, 19, 8, 18, 7, 17, 6, 16, 5, 15, 4, 14, 3, 13, 2, 12, 1, 11	20	32
FC 4959	10, 20, 9, 19, 8, 18, 7, 17, 6, 16, 5, 15, 4, 14, 3, 13, 2, 12, 1, 11	20	20
FC 4962	10, 20, 9, 19, 8, 18, 7, 17, 6, 16, 5, 15, 4, 14, 3, 13, 2, 12, 1, 11	20	192 for Model 595 160 for Model 590
FC 5723	10, 20, 9, 19, 8, 18, 7, 17, 6, 16, 5, 15, 4, 14, 3, 13, 2, 12, 1, 11	20	32
FC 6204	10, 20, 9, 19, 8, 18, 7, 17, 6, 16, 5, 15, 4, 14, 3, 13, 2, 12, 1, 11	20	32

** Extra High Bandwidth (EHB) adapter. See the Performance notes before installing this adapter.
* High Bandwidth (HB) adapter. See the Performance notes before installing this adapter.

C

Installation planning

Planning for model 9119-590 and 9119-595 server specifications This topic gives you a thorough understanding of the model 9119-590, and 9119-595 server specifications, including dimensions, electrical, power, temperature, environment, and service clearances. You will also find links to more detailed information, such as compatible hardware and plug types. The IBM ⓔserver p5 590 and p5 595 consist of multiple components, as summarized in the following table.

Doors and covers

Doors and covers are an integral part of the system and are required for product safety and EMC compliance. The following rear door options are available for the model 9119-590 and 9119-595:

Enhanced acoustical cover option

This feature provides a low-noise option for clients or sites with stringent acoustical requirements and where a minimal system footprint is not critical. The acoustical cover option consists of a special front and rear doors that are approximately 250 mm (10 in.) deep and contain acoustical treatment that lowers the noise level of the machine by approximately 7 dB (0.7 B) compared to the Slimline doors. This reduction in noise emission levels means that the noise level of a single model system with Slimline covers is about the same as the noise level of five model systems with acoustical covers.

Slimline cover option

This feature provides a smaller-footprint and lower-cost option for clients or sites where space is more critical than acoustical noise levels. The Slimline cover option consists of a front door, which is approximately 100 mm (4 in.) deep, and a rear door, which is approximately 50 mm (2 in.) deep. No acoustical treatment is available for this option.

Raised-floor requirements and preparation

A raised-floor is required for the model 9119-595 and its associated racks to ensure optimal performance and to comply with electromagnetic compatibility (EMC) requirements. A raised-floor is not required for the model 9119-590, but it is recommended for optimum system cooling and cable management. Raised-floor cutouts should be protected by electrically nonconductive molding, appropriately sized, with edges treated to prevent cable damage and to prevent casters from rolling into the floor cutouts.

Front-service access is necessary on the model 9119-590 and 9119-595 to accommodate a lift tool for the servicing of large drawers (the processor books and I/O drawers). Front and rear service access is necessary to accommodate the lift tool for servicing of the optional integrated battery backup (IBB).

Securing the rack

The following can be ordered by the client as additional rack-securing options for the model 9119-590 and 9119-595.

- RPQ 8A1183 for attaching the rack-mounting plates to the concrete floor (non-raised floor)
- RPQ 8A1185 to attach the rack to a concrete floor when on a raised floor (9 1/2 inches to 11 3/4 inches high)
- RPQ 8A1186 to attach the rack to a concrete floor when on a raised floor (11 3/4 inches to 16 inches high)

Considerations for multiple system installations

In a multi-frame installation, it is possible that a floor tile with cable cutouts will bear two concentrated static loads up to 476 kg (1050 lb.) per caster/leveler. Thus, the total concentrated load can be as high as 953 kg (2100 lb.). Contact the floor tile manufacturer or consult a structural engineer to ensure that the raised floor assembly can support this load.

When you are integrating a model 9119-590 and 9119-595 into an existing multiple-system environment, or when adding additional systems to an installed 9119-590 and 9119-595, consider the following factors:

Minimum aisle width

For multiple rows of systems containing one or more model 9119-590 or 9119-595, the minimum aisle width in the front of the system is 1118 mm (44 in.) and 1041 mm (33 in.) in the rear of the system to allow room to perform service operations. The minimum aisle width is in addition to the front and rear service clearances of 1219 mm (48 in.) and 914 mm (36 in.), respectively. Service clearances are measured from the edges of the frame (with doors open) to the nearest obstacle.

Thermal interactions

Systems should be faced front-to-front and rear-to-rear to create •cool• and •hot• aisles to maintain effective system thermal conditions.

Cool aisles need to be of sufficient width to support the airflow requirements of the installed systems. The airflow per tile will be dependent on the underfloor pressure and perforations in the tile. A typical underfloor pressure of 0.025 in. of water will supply 300-400 cfm through a 25 percent open 2 ft. by 2 ft. floor tile.

Moving the system to the installation site

You should determine the path that must be taken to move the system from the delivery location to the installation site. You should verify that the height of all doorways, elevators, and so on are sufficient to allow moving the system to the installation site. You should also verify that the weight limitations of elevators, ramps, floors, floor tiles, and so on are sufficient to allow moving the system to the installation site. If the height or weight of the system can cause a problem when the system is moved to the installation site, you should contact your local site planning, marketing, or sales representative.

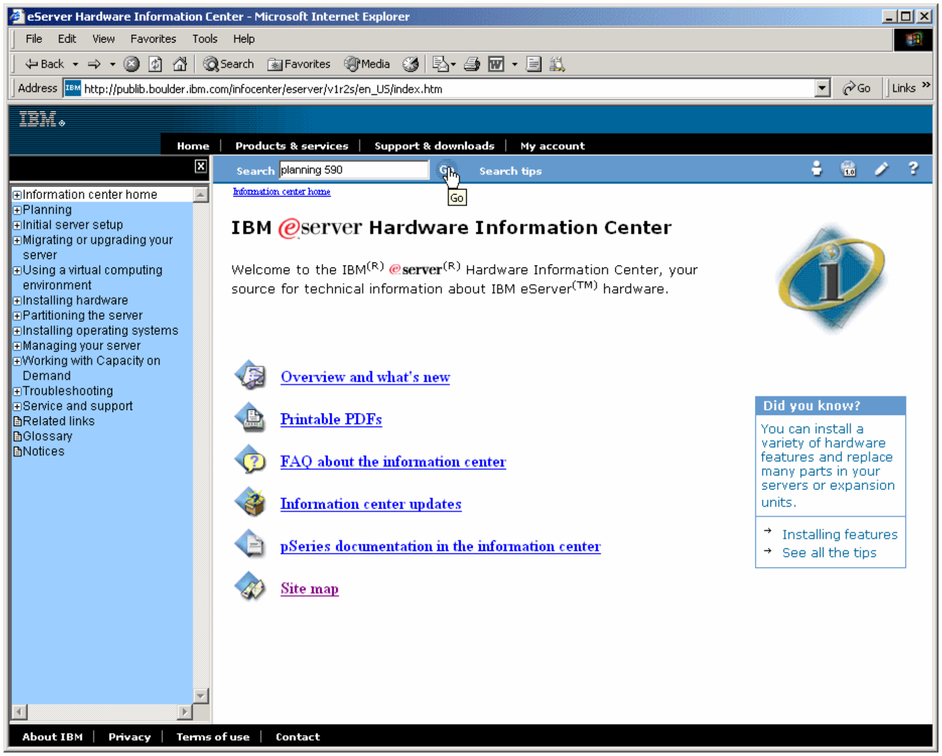
Dual power installation

Some model 9119-590 and 9119-595 configurations are designed with a fully redundant power system. These systems have two power cords attached to two power input ports which, in turn, power a fully redundant power distribution system within the system. To take full advantage of the redundancy/reliability that is built into the computer system, the system must be powered from two distribution panels. Larger model 9119-590 and 9119-595 configurations require power from two power cords, and they do not have redundant power cords. The possible power installation configurations are described as follows. See Dual power installations for additional information about power.

Planning and installation documentation

For all information about planning and installation go to the IBM @server Hardware Information Center and search for ‘installation planning 595’. The following is provided as an example:

http://publib.boulder.ibm.com/infocenter/eserver/v1r2s/en_US/index.htm


eServer Hardware Information Center - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://publib.boulder.ibm.com/infocenter/eserver/v1r2s/en_US/index.htm
Go Links
IBM
Home | Products & services | Support & downloads | My account
Search planning 590 Go Search tips
Information center home
Planning
Initial server setup
Migrating or upgrading your server
Using a virtual computing environment
Installing hardware
Partitioning the server
Installing operating systems
Managing your server
Working with Capacity on Demand
Troubleshooting
Service and support
Related links
Glossary
Notices
Overview and what's new
Printable PDFs
FAQ about the information center
Information center updates
pSeries documentation in the information center
Site map
Installing features
See all the tips
About IBM | Privacy | Terms of use | Contact

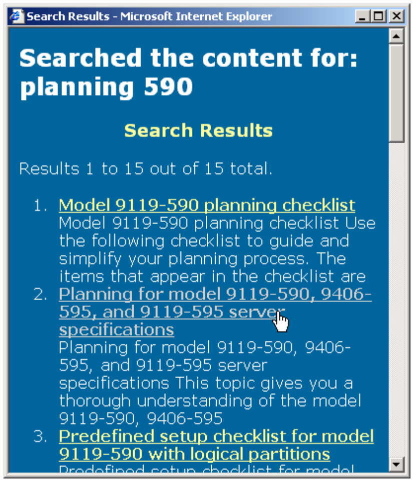
IBM @server Hardware Information Center

Welcome to the IBM(R) @server(R) Hardware Information Center, your source for technical information about IBM eServer(TM) hardware.

Did you know?

You can install a variety of hardware features and replace many parts in your servers or expansion units.

Figure C-1 Search for planning



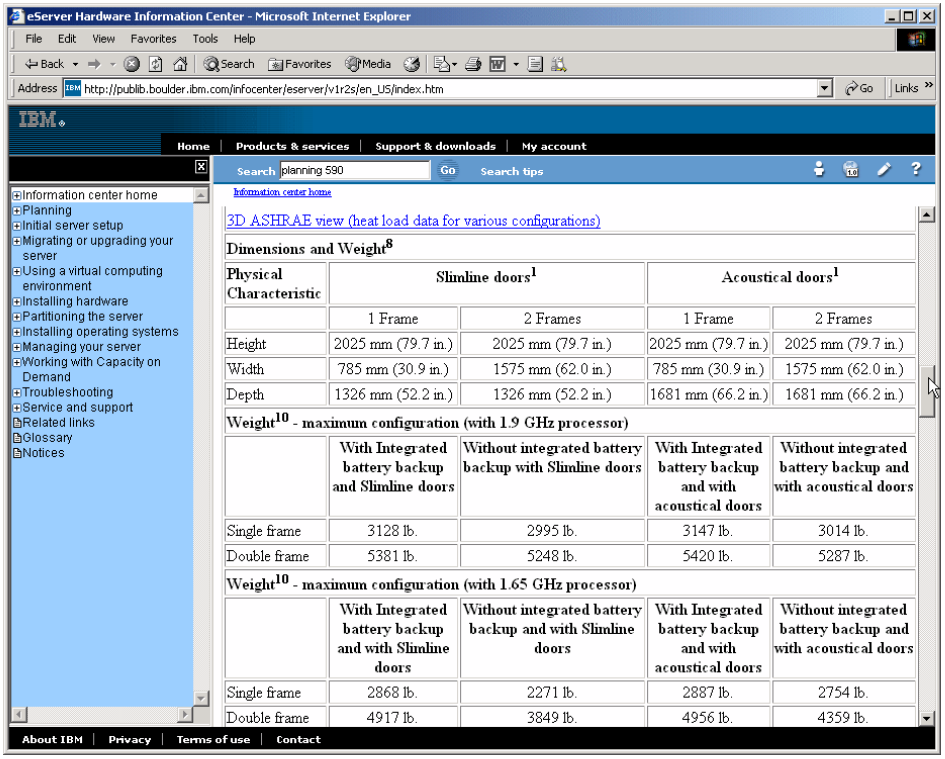
Searched the content for: planning 590

Search Results

Results 1 to 15 out of 15 total.

1. Model 9119-590 planning checklist
Model 9119-590 planning checklist Use the following checklist to guide and simplify your planning process. The items that appear in the checklist are
2. Planning for model 9119-590, 9406-595, and 9119-595 server specifications
Planning for model 9119-590, 9406-595, and 9119-595 server specifications This topic gives you a thorough understanding of the model 9119-590, 9406-595
3. Predefined setup checklist for model 9119-590 with logical partitions
Predefined setup checklist for model

Figure C-2 Select 9119-590 and 9119-595


eServer Hardware Information Center - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://publib.boulder.ibm.com/infocenter/eserver/v1r2s/en_US/index.htm
Go Links
IBM
Home | Products & services | Support & downloads | My account
Search planning 590 Go Search tips
Information center home
Planning
Initial server setup
Migrating or upgrading your server
Using a virtual computing environment
Installing hardware
Partitioning the server
Installing operating systems
Managing your server
Working with Capacity on Demand
Troubleshooting
Service and support
Related links
Glossary
Notices
Information center home
About IBM | Privacy | Terms of use | Contact

3D ASHRAE view (heat load data for various configurations)

Dimensions and Weight[8]

Physical Characteristic	Slimline doors[1]		Acoustical doors[1]	
	1 Frame	2 Frames	1 Frame	2 Frames
Height	2025 mm (79.7 in.)	2025 mm (79.7 in.)	2025 mm (79.7 in.)	2025 mm (79.7 in.)
Width	785 mm (30.9 in.)	1575 mm (62.0 in.)	785 mm (30.9 in.)	1575 mm (62.0 in.)
Depth	1326 mm (52.2 in.)	1326 mm (52.2 in.)	1681 mm (66.2 in.)	1681 mm (66.2 in.)
Weight[10] - maximum configuration (with 1.9 GHz processor)				
	With Integrated battery backup and Slimline doors	Without integrated battery backup with Slimline doors	With Integrated battery backup and with acoustical doors	Without integrated battery backup and with acoustical doors
Single frame	3128 lb.	2995 lb.	3147 lb.	3014 lb.
Double frame	5381 lb.	5248 lb.	5420 lb.	5287 lb.
Weight[10] - maximum configuration (with 1.65 GHz processor)				
	With Integrated battery backup and with Slimline doors	Without integrated battery backup and with Slimline doors	With Integrated battery backup and with acoustical doors	Without integrated battery backup and with acoustical doors
Single frame	2868 lb.	2271 lb.	2887 lb.	2754 lb.
Double frame	4917 lb.	3849 lb.	4956 lb.	4359 lb.

Figure C-3 Planning information

D

System documentation

The system documentation is installed on the HMC and will be updated during the update of the HMC. The updates are found at the following Web site:

http://techsupport.services.ibm.com/server/hmc/power5

Under Version 4.2 machine code updates you will find the updates for the HMC and also updates for the InfoCenter of the HMC, using the following URL.

http://techsupport.services.ibm.com/server/hmc/power5/fixes/v4r2.html

Firmware (LIC) download is available at the following URL:

http://techsupport.services.ibm.com/server/mdownload2/download.html

If you plan to update more than the service processor (LIC) you should download the CD-ROM image. This CD can be mounted on the HMC and used to update all hardware devices.

http://techsupport.services.ibm.com/server/mdownload2/cdimage.html

IBM @server Hardware Information Center

The following describes the interface used to obtain reference information for IBM servers, management consoles, and other products.

What is the Hardware Information Center?

The following are the primary characteristics of the hardware Information Center

- Web-Based information
- Contains info for all i5, p5, and OpenPower hardware models
- Planning
- Installing hardware
- Partitioning
- Capacity on Demand
- Managing the server using the HMC or other consoles
- Customer troubles shooting and Services
- Service provider (CE, SSR, PE) service instructions
- Contains pointers to i5/OS, AIX, Linux Info Centers or Web sites for software information

How do I get it?

Obtaining the Information Center is through one of several ways.

On the Internet

Use one of the following Web sites to find the Information Center that is correct for you.

```
http://publib.boulder.ibm.com/eserver
http://publib.boulder.ibm.com/infocenter/eserver/v1r2s/en_US/index.htm
```

On an HMC

Click on Information Center and Setup Wizard

On a CD-ROM

Shipped with the hardware (SK3T-8159)

As an order from the IBM Publications Center

You can order the Information Center through the IBM publication ordering center using the following URL.

`http://www.ibm.com/shop/publications/order`

To use the CD you receive with this order, follow the these steps:

- Install to your PC, laptop, LAB drive or a partition on your system
- If autorun is set on your operating system, the installation program starts when you insert the CD in the CD-ROM drive.
- If you do not have autorun set, you must run the appropriate install program on the CD:
- For Windows 2000, ME, XP, NT 4.0, or i5/OS run win32/install_win.exe
- For SuSE or Red Hat on Intel processor run linux_i386/install_lnx_i386.bin
- For SuSE or Red Hat on POWER5 processor run linux_ppc/install_lnx_ppc.bin
- For AIX run aix/install_aix.bin

How do I get updates?

There are two major ways to obtain updates to the Information Center, as in the following.

Subscription

One method is through a subscription.

The eserver Hardware Information Center is updated on the Internet for the following situations:

- New hardware or function is released. The Web site is updated on the Announce date.
- Severity 2 issues or problems within the information. The Web site is updated at month-end.
- Severity 1 problem within the information. The Web site is updated as soon as possible.

Subscribe to receive e-mail notifications when updates are made.

- From the Info Center home page, click on Informing Center updates HMC fixes.

  This will cause the latest version of the Information Center to be updated when you download fixes for your HMC.

Download updates from the Internet

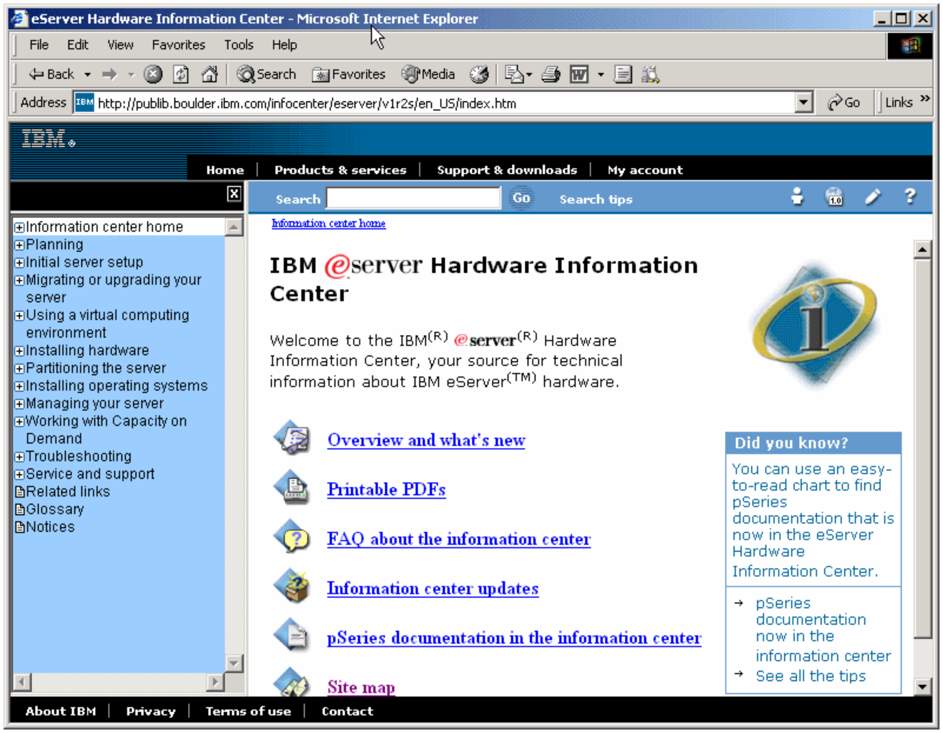
If you have used the CD-ROM to install the information center, you can download updates from the Internet to keep your locally installed version updated.

From your PC, follow these steps

- Select the Windows Start Menu
- Locate in the installed programs list IBM @server Hardware Information Center
- In this menu, locate Configuration Tools
- Select Download Updates from the Internet

How do I use the application?

Opening the information center results in a panel similar to Figure D-1.


eServer Hardware Information Center - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address
Go Links
IBM.
Home | Products & services | Support & downloads | My account
Search Go Search tips
Information center home
Information center home
Planning
Initial server setup
Migrating or upgrading your server
Using a virtual computing environment
Installing hardware
Partitioning the server
Installing operating systems
Managing your server
Working with Capacity on Demand
Troubleshooting
Service and support
Related links
Glossary
Notices
About IBM | Privacy | Terms of use | Contact

http://publib.boulder.ibm.com/infocenter/eserver/v1r2s/en_US/index.htm

IBM @server Hardware Information Center

Welcome to the IBM(R) @server(R) Hardware Information Center, your source for technical information about IBM eServer(TM) hardware.

Overview and what's new

Printable PDFs

FAQ about the information center

Information center updates

pSeries documentation in the information center

Site map

Did you know?

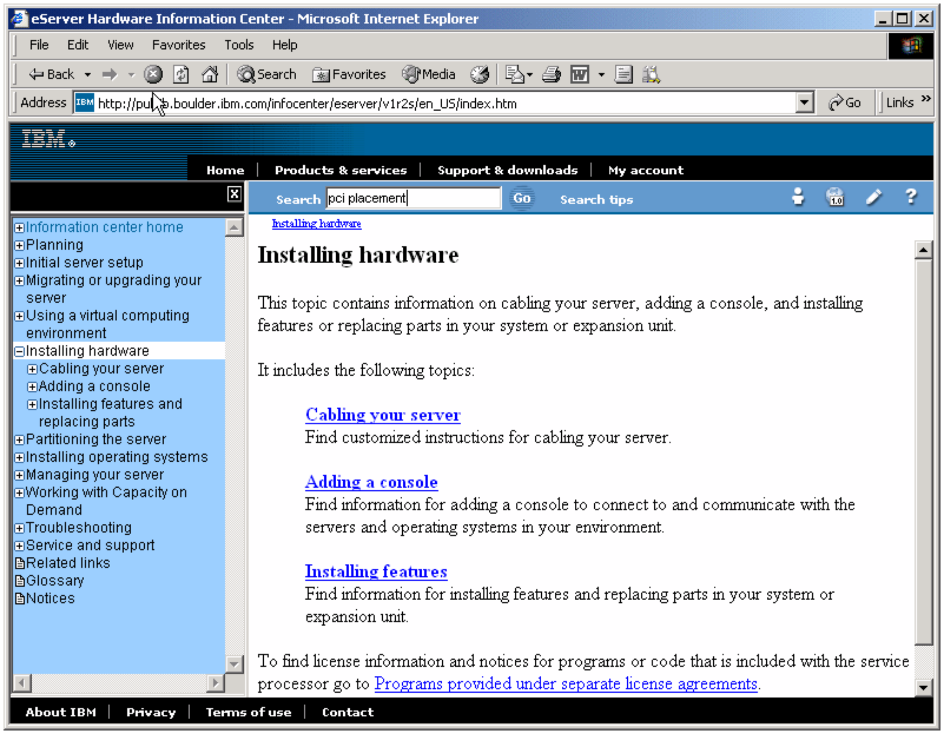
You can use an easy-to-read chart to find pSeries documentation that is now in the eServer Hardware Information Center.

→ pSeries documentation now in the information center

→ See all the tips

Figure D-1 Information Center

Full-text search is available on the search term that you type in the search field (Figure D-2).


eServer Hardware Information Center - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://publib.boulder.ibm.com/infocenter/eserver/v1r2s/en_US/index.htm
Go Links
IBM
Home | Products & services | Support & downloads | My account
Search pci placement Go Search tips
About IBM | Privacy | Terms of use | Contact

- Information center home
- Planning
- Initial server setup
- Migrating or upgrading your server
- Using a virtual computing environment
- Installing hardware
  - Cabling your server
  - Adding a console
  - Installing features and replacing parts
- Partitioning the server
- Installing operating systems
- Managing your server
- Working with Capacity on Demand
- Troubleshooting
- Service and support
- Related links
- Glossary
- Notices

Installing hardware

Installing hardware

This topic contains information on cabling your server, adding a console, and installing features or replacing parts in your system or expansion unit.

It includes the following topics:

Cabling your server
Find customized instructions for cabling your server.

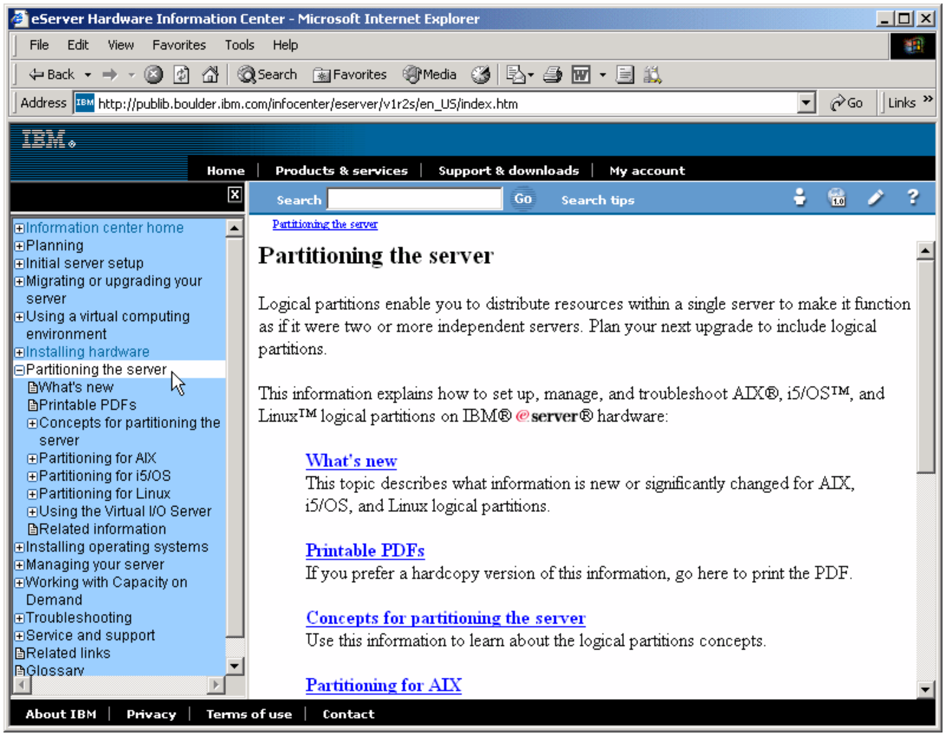
Adding a console
Find information for adding a console to connect to and communicate with the servers and operating systems in your environment.

Installing features
Find information for installing features and replacing parts in your system or expansion unit.

To find license information and notices for programs or code that is included with the service processor go to Programs provided under separate license agreements.

Figure D-2 Search field

The Navigation bar is in the left frame. List all the categories and topics (Figure D-3).


eServer Hardware Information Center - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Go Links
IBM
Home | Products & services | Support & downloads | My account
Search Go Search tips
About IBM | Privacy | Terms of use | Contact

Address http://publib.boulder.ibm.com/infocenter/eserver/v1r2s/en_US/index.htm

- ⊞Information center home
- ⊞Planning
- ⊞Initial server setup
- ⊞Migrating or upgrading your server
- ⊞Using a virtual computing environment
- ⊞Installing hardware
- ⊟Partitioning the server
  - What's new
  - Printable PDFs
  - ⊞Concepts for partitioning the server
  - ⊞Partitioning for AIX
  - ⊞Partitioning for i5/OS
  - ⊞Partitioning for Linux
  - ⊞Using the Virtual I/O Server
  - Related information
- ⊞Installing operating systems
- ⊞Managing your server
- ⊞Working with Capacity on Demand
- ⊞Troubleshooting
- ⊞Service and support
- Related links
- Glossary

Partitioning the server

Partitioning the server

Logical partitions enable you to distribute resources within a single server to make it function as if it were two or more independent servers. Plan your next upgrade to include logical partitions.

This information explains how to set up, manage, and troubleshoot AIX®, i5/OS™, and Linux™ logical partitions on IBM® @server® hardware:

What's new
This topic describes what information is new or significantly changed for AIX, i5/OS, and Linux logical partitions.

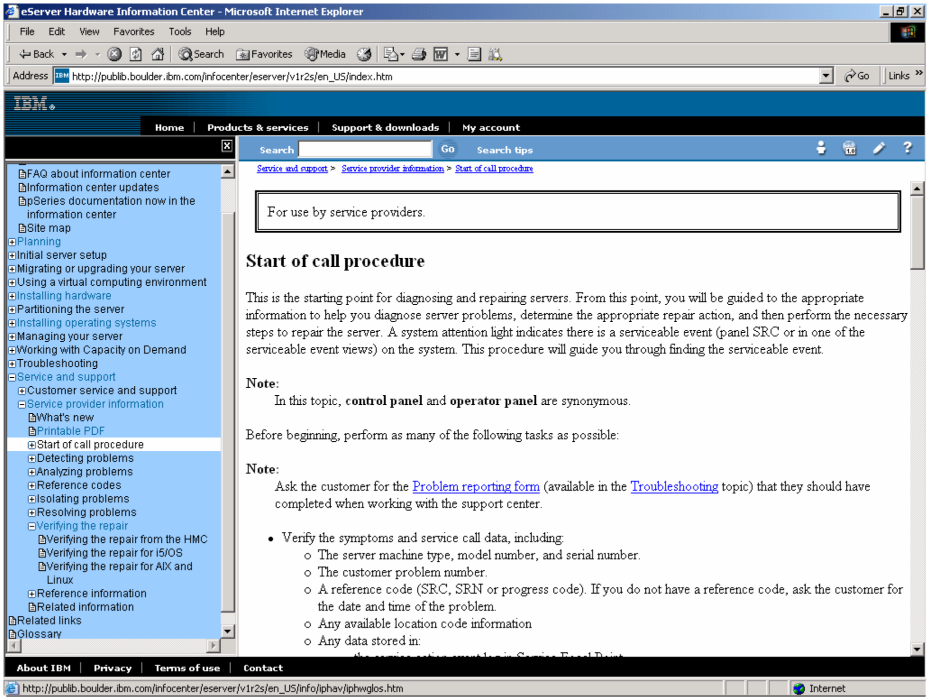
Printable PDFs
If you prefer a hardcopy version of this information, go here to print the PDF.

Concepts for partitioning the server
Use this information to learn about the logical partitions concepts.

Partitioning for AIX

Figure D-3 Navigation bar

In the toolbar we can customizing the Interface, selecting another language, sending feedback, or getting help. It is located in the upper right corner (Figure D-4).


eServer Hardware Information Center - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://publib.boulder.ibm.com/infocenter/eserver/v1r2s/en_US/index.htm
Go Links
IBM.
Home | Products & services | Support & downloads | My account
Search Go Search tips
FAQ about information center
Information center updates
pSeries documentation now in the information center
Site map
Planning
Initial server setup
Migrating or upgrading your server
Using a virtual computing environment
Installing hardware
Partitioning the server
Installing operating systems
Managing your server
Working with Capacity on Demand
Troubleshooting
Service and support
Customer service and support
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 and Linux
Reference information
Related information
Related links
Glossary
Service and support > Service provider information > Start of call procedure
About IBM | Privacy | Terms of use | Contact
http://publib.boulder.ibm.com/infocenter/eserver/v1r2s/en_US/info/iphav/iphwglos.htm
Internet

For use by service providers.

Start of call procedure

This is the starting point for diagnosing and repairing servers. From this point, you will be guided to the appropriate information to help you diagnose server problems, determine the appropriate repair action, and then perform the necessary steps to repair the server. A system attention light indicates there is a serviceable event (panel SRC or in one of the serviceable event views) on the system. This procedure will guide you through finding the serviceable event.

Note:
In this topic, **control panel** and **operator panel** are synonymous.

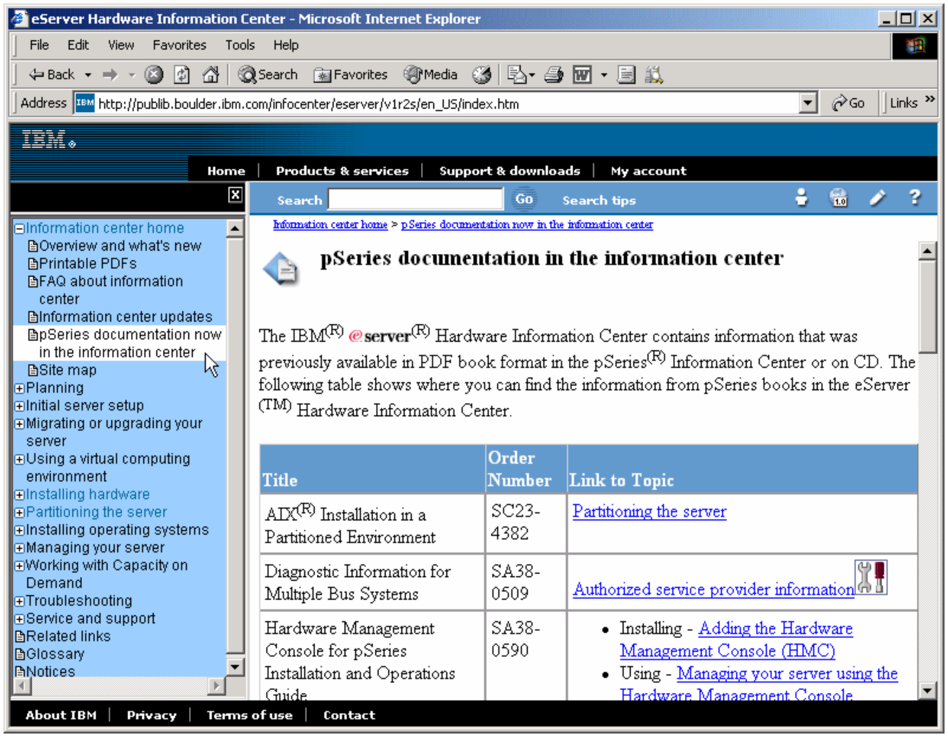
Before beginning, perform as many of the following tasks as possible:

Note:
Ask the customer for the Problem reporting form (available in the Troubleshooting topic) that they should have completed when working with the support center.

- Verify the symptoms and service call data, including:
  - The server machine type, model number, and serial number.
  - The customer problem number.
  - A reference code (SRC, SRN or progress code). If you do not have a reference code, ask the customer for the date and time of the problem.
  - Any available location code information
  - Any data stored in:
    - the service action event log in Service Focal Point

Figure D-4 Toolbar with start off call

You can also find previous pSeries publications in the Information Center. If you search for order number within the page, click on the right frame and press Ctrl+F (Figure D-5).


eServer Hardware Information Center - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://publib.boulder.ibm.com/infocenter/eserver/v1r2s/en_US/index.htm Go Links
IBM
Home | Products & services | Support & downloads | My account
Search Go Search tips
Information center home
Overview and what's new
Printable PDFs
FAQ about information center
Information center updates
pSeries documentation now in the information center
Site map
Planning
Initial server setup
Migrating or upgrading your server
Using a virtual computing environment
Installing hardware
Partitioning the server
Installing operating systems
Managing your server
Working with Capacity on Demand
Troubleshooting
Service and support
Related links
Glossary
Notices
Information center home > pSeries documentation now in the information center
About IBM | Privacy | Terms of use | Contact

pSeries documentation in the information center

The IBM® @server® Hardware Information Center contains information that was previously available in PDF book format in the pSeries® Information Center or on CD. The following table shows where you can find the information from pSeries books in the eServer (TM) Hardware Information Center.

Title	Order Number	Link to Topic
AIX® Installation in a Partitioned Environment	SC23-4382	Partitioning the server
Diagnostic Information for Multiple Bus Systems	SA38-0509	Authorized service provider information
Hardware Management Console for pSeries Installation and Operations Guide	SA38-0590	• Installing - Adding the Hardware Management Console (HMC) • Using - Managing your server using the Hardware Management Console

Figure D-5 Previous pSeries documentation

If you has comments to InfoCenter feel free to sent us and use the Feedback button on the upper right side.

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 on ordering these publications, see “How to get IBM Redbooks” on page 275. Note that some of the documents referenced here may be available in softcopy only.

- IBM @server pSeries 670 and pSeries 690 System Handbook, SG24-7040
- *Advanced POWER Virtualization on IBM p5 @server Introduction and Basic Configuration, SG24-7940*
- *AIX 5L Differences Guide, SG24-5765*
- *Managing AIX Server Farms, SG24-6606*
- *IBM Hardware Management Console for pSeries Maintenance Guide,*
- *SA38-0603*
- *Installation Guide 61D I/O drawer 61R Second I/O Rack, SA23-1281*
- *RS/6000 and @server pSeries Diagnostics Information for Multiple Bus Systems, SA38-0509*
- *IBM @server pSeries 690 Service Guide, SA38-0589*

Other publications

These publications are also relevant as further information sources:

- IBM @server pSeries 690 Availability Best Practices
- The IBM @server pSeries 690 Reliability, Availability, Serviceability (RAS)
- AIX 5L Version 5.2 AIX Installation in a Partitioned Environment
- AIX 5L Version 5.3 AIX Installation in a Partitioned Environment
- AIX 5L Version 5.2 Installation Guide and Reference
- AIX 5L Version 5.3 Installation Guide and Reference
- AIX 5L Version 5.2 Reference Documentation: Commands Reference

- AIX 5L Version 5.2 System Management Guide: Communications and Networks
- AIX 5L Version 5.2 System Management Guide: Operating System and Devices

The following whitepapers can be found on the Internet:

- *IBM @server p5 Introduction to the Virtual I/O Server Whitepaper*

  http://www-1.ibm.com/servers/eserver/pseries/hardware/whitepapers/virtual_io.pdf
- *LPAR for Decision Makers Whitepaper*

  http://www-1.ibm.com/servers/eserver/pseries/hardware/whitepapers/lpar_decision.pdf
- *Linux Network Installation Service for multiple platforms provided by an AIX 5L Version 5.3 NIM Server Whitepaper*

  http://www-1.ibm.com/servers/eserver/pseries/hardware/whitepapers/network_install.pdf
- *Virtual Networking on AIX 5L Whitepaper*

  http://www-1.ibm.com/servers/aix/whitepapers/aix_vn.pdf
- *IBM @server p5 - AIX 5L Support for Micro-Partitioning and Simultaneous Multi-threading Whitepaper*

  http://www-1.ibm.com/servers/aix/whitepapers/aix_support.pdf
- *IBM @server Linux on POWER Overview Whitepaper*

  http://www-1.ibm.com/servers/eserver/linux/power/whitepapers/linux_overview.pdf

Online resources

These Web sites and URLs are also relevant as further information sources:

- IBM @server Information Center

  http://publib.boulder.ibm.com/eserver/
- AIX 5L operating system and related IBM products information

  http://www.ibm.com/servers/aix/
- AIX toolkit for Linux applications

  http://www.ibm.com/servers/aix/products/aixos/linux/download.html

- Application availability on the AIX 5L operating system (alphabetical listing and advanced search options for IBM software products and third-party software products)

  http://www.ibm.com/servers/aix/products/
- Capacity on Demand (CoD) offering summary

  http://www.ibm.com/servers/eserver/pseries/ondemand/cod/
- IBM AIX 5L Solution Developer Application Availability Web page

  http://www.ibm.com/servers/aix/isv/availability.html
- IBM AIX: IBM Application Availability Guide Web page

  http://www.ibm.com/servers/aix/products/ibmsw/list
- Linux on pSeries information

  http://www.ibm.com/servers/eserver/pseries/linux/
- CoD activation Web site

  http://www-912.ibm.com/pod/pod

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

Abbreviations and acronyms

AC	Alternating Current
AIO	Asynchronous I/O
AIX	Advanced Interactive Executive
APAR	Authorized Program Analysis Report
ARP	Address Resolution Protocol
ASMI	Advanced System Management Interface
ATM	Asynchronous Transfer Mode
ATS	Advanced Technical Support
BIST	Built-In Self-Test
BPA	Bulk Power Adapter
BPC	Bulk Power Controller
BPD	Bulk Power Distribution
BPR	Bulk Power Regulators
CD	Compact Disk
CD-R	CD Recordable
CD-ROM	Compact Disk-Read Only Memory
CE	Customer Engineer
CEC	Central Electronics Complex
CHRP	Common Hardware Reference Platform
CMOS	Complimentary Metal-Oxide Semiconductor
CoD	Capacity on Demand
CPU	Central Processing Unit
CRC	Cyclic Redundancy Check
CRT	Cathode Ray Tube
CSM	Cluster Systems Management
CSR	Customer Service Representative
CSU	Customer Set-Up
CUoD	Capacity Upgrade on Demand
CWS	Control Workstation
DASD	Direct Access Storage Device
DC	Direct Current
DCA	Distributed Converter Assembly
DDR	Double Data Rate
DGD	Dead Gateway Detection
DHCP	Dynamic Host Configuration Protocol
DIMM	Dual In-Line Memory Module
DLPAR	Dynamic LPAR
DMA	Direct Memory Access
DNS	Domain Naming System
DOS	Disk Operating System
DPCL	Dynamic Probe Class Library
DRAM	Dynamic Random Access Memory
DVD	Digital Versatile Disk
EC	Engineering Change
ECC	Error Checking and Correcting
EEH	Extended Error Handling
EHB	Extra High Bandwidth
EEPROM	Electrically Erasable Programmable Read Only Memory
EIA	Electronic Industries Association
ELA	Error Log Analysis
EMC	Electromagnetic Compatibility

EPOW	Environmental and Power Warning
ESA	Electronic Service Agreement
F/C	Feature Code
FC	Fibre Channel
FFDC	First Failure Data Capture
FIR	Fault Isolation Register
FLASH EPROM	Flash Erasable Programmable Read-Only Memory
FPR	Floating Point Register
FPU	Floating Point Unit
GB	Gigabyte
GPR	General-Purpose Register
GUI	Graphical User Interface
GX	Gigabit Bus
HACMP	High Availability Cluster Multi Processing
HMC	Hardware Management Console
HPS	High Performance Supercomputer
HPT	Hardware Page Table
I/O	Input/Output
IBB	Internal Battery Backup
IBF	Internal Battery Feature
IBM	International Business Machines
ICMP	Internet Control Message Protocol
ID	Identification
IEEE	Institute of Electrical and Electronics Engineers
IP	Internetwork Protocol (OSI)
IPL	Initial Program Load
ISV	Independent Software Vendor
ITS	International Technical Support
ITSO	International Technical Support Organization
L1	Level 1
L2	Level 2
L3	Level 3
LACP	Link Aggregation Control Protocol
LAN	Local Area Network
LED	Light Emitting Diode
LIC	Licensed Internal Code
LMB	Logical Memory Block
LPAR	Logical Partition
LVM	Logical Volume Manager
LVT	LPAR Validation Tool
MAC	Media Access Control
Mbps	Megabits Per Second
MBps	Megabytes Per Second
MCM	Multichip Module
MES	Miscellaneous Equipment Specification
MTU	Maximum Transmission Unit
NAT	Network Address Translation
NDP	Neighbor Discovery Protocol
NIM	Network Installation Management
NVRAM	Non-Volatile Random Access Memory
OEM	Original Equipment Manufacturer
PC	Personal Computer
PCI	Peripheral Component Interconnect
PCI	PCI Extended
PHB	Processor Host Bridges
PHYP	Power Hypervisor
PLM	Partition Load Manager
POST	Power-On Self-test

POWER	Performance Optimization with Enhanced Risc (Architecture)
PPP	Point-to-Point Protocol
PRU	Processor Unit Value
PTF	Program Temporary Fix
PVID	Physical Volume Identifier
RAID	Redundant Array of Independent Disks
RAS	Reliability, Availability, and Serviceability
RIO	Remote I/O
RPM	Redhat Package Manager
RPQ	Request for Price Quote
RTAS	Runtime Abstraction Services
RSE	Register Stack Engine
SAG	Service Agent Gateway
SCSI	Small Computer System Interface
SEA	Shared Ethernet Adapter
SES	SCSI Enclosure Services
SFP	Service Focal Point
SMI	Sync. Memory Interface
SMP	Symmetric Multiprocessor
SMS	System Management Services
SMTP	Simple Mail Transport Protocol
SOD	Statement of Direction
SOI	Silicon-on-Insulator
SP	IBM RS/6000 Scalable POWER parallel Systems
SP	Service Processor
SPCN	System Power Control Network
SSA	Serial Storage Architecture
SSH	Secure Shell
SSL	Secure Socket Layer
SSR	System Service Representative
SWMA	Software Maintenance Agreement
SYNC	Synchronization
TCO	Total Cost of Ownership
TCP/IP	Transmission Control Protocol/Internet Protocol
TPO	Timed Power On
UDP	User Datagram Protocol
UPS	Uninterrupable Power Supply
URL	Uniform Resource Locator
USB	Universal Serial Bus
VIO	Virtual I/O
VIOS	Virtual I/O Server
VLAN	Virtual Local Area Network
VMM	Virtual Memory Manager
VP	Virtual Processor
VPD	Vital Product Data
VPN	Virtual Private Network
VSCSI	Virtual SCSI
WSMRC	Web-based System Management Remote Clients

Index

To determine the spine width of a book, you divide the paper PPI into the number of pages in the book. An example is a 250 page book using Plainfield opaque 50# smooth which has a PPI of 526. Divided 250 by 526 which equals a spine width of .4752". In this case, you would use the .5" spine. Now select the Spine width for the book and hide the others: **Special>Conditional Text>Show/Hide>SpineSize(-->Hide:)>Set** . Move the changed Conditional text settings to all files in your book by opening the book file with the spine.fm still open and **File>Import>Formats** the Conditional Text Settings (ONLY!) to the book files.

Redbooks

IBM server p5 590 and p5 595 System Handbook

(0.5” spine)
0.475”<->0.875”
250 <-> 459 pages

To determine the spine width of a book, you divide the paper PPI into the number of pages in the book. An example is a 250 page book using Plainfield opaque 50# smooth which has a PPI of 526. Divided 250 by 526 which equals a spine width of .4752". In this case, you would use the .5" spine. Now select the Spine width for the book and hide the others: **Special>Conditional Text>Show/Hide>SpineSize(-->Hide:)>Set** . Move the changed Conditional text settings to all files in your book by opening the book file with the spine.fm still open and **File>Import>Formats** the Conditional Text Settings (ONLY!) to the book files.

IBM eserver p5 590 and p5 595 System Handbook

Component-based description of the hardware architecture

A guide for machine type 9119 models 590 and 595

Capacity on Demand explained

This IBM Redbook explains the IBM eserver p5 models 590 and 595 (9119-590, 9119-595), a new level of UNIX servers providing world-class performance, availability, and flexibility. Ideal for on demand computing environments, data center implementation, application service providers, and high performance computing, this new class of high-end servers include mainframe-inspired self-management and security to meet your most demanding needs. The IBM eserver p5 590 and p5 595 provide an expandable, high-end enterprise solution for managing the computing requirements needed to become an on demand business.

This publication includes the following topics:

Overview of the p5-590 and p5-595
Hardware architecture of the p5-590 and p5-595
Overview of virtualization features
Capacity on Demand overview
Reliability, availability, and serviceability
Hardware Management Console features and functions

This publication is an ideal desk-side reference for IBM professionals, Business Partners, and technical specialists who support the p5-590 and p5-595, and for those who want to learn more about this radically new server in a clear, single-source handbook.

INTERNATIONAL TECHNICAL SUPPORT ORGANIZATION

BUILDING TECHNICAL INFORMATION BASED ON PRACTICAL EXPERIENCE

SG24-9119-00 ISBN