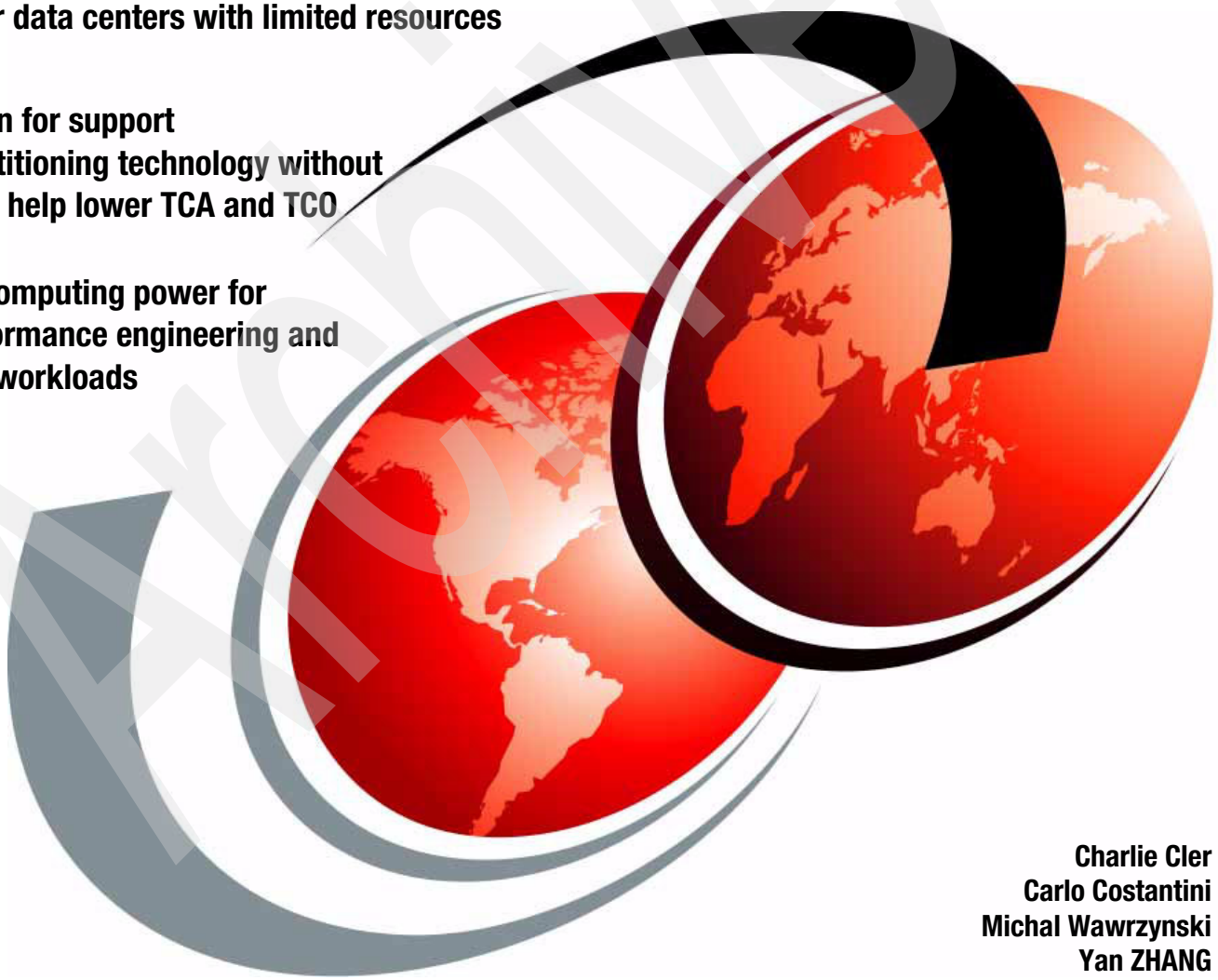
IBM

IBM System p5 505 and 505Q
Technical Overview and Introduction

- High-performance server in a dense, 1U package is ideal for data centers with limited resources
- New option for support Micro-Partitioning technology without an HMC to help lower TCA and TCO
- The raw computing power for high-performance engineering and scientific workloads



Charlie Cler
Carlo Costantini
Michal Wawrzynski
Yan ZHANG

Redpaper

ibm.com/redbooks

International Technical Support Organization

IBM System p5 505 and 505Q Technical Overview and Introduction

September 2006

Note: Before using this information and the product it supports, read the information in “Notices” on page vii.

Second Edition (September 2006)

This edition applies to IBM System p5 505 and 505Q (product number 9115-505), Linux, and IBM AIX 5L Version 5.3, product number 5765-G03.

Contents

Notices

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

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

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

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, 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:

@server®	HACMP™	Redbooks™
Redbooks (logo)  ™	IBM®	Resource Link™
eServer™	Micro-Partitioning™	RS/6000®
pSeries®	OpenPower™	Service Director™
AIX 5L™	PowerPC®	System p™
AIX®	POWER™	System p5™
Chipkill™	POWER Hypervisor™	System Storage™
DS4000™	POWER4™	TotalStorage®
DS6000™	POWER5™	Virtualization Engine™
DS8000™	POWER5+™	1350™
Enterprise Storage Server®	PTX®	

The following terms are trademarks of other companies:

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

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

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

Preface

This IBM® Redpaper is a comprehensive guide covering the IBM System p5™ 505 and 505Q server supporting the IBM AIX® 5L™ and Linux® operating systems. We introduce major hardware offerings and discuss their prominent functions.

Professionals wanting to acquire a better understanding of IBM System p5 products should consider reading this document. The intended audience includes:

- Clients
- Marketing representatives
- Technical support professionals
- IBM Business Partners
- Independent software vendors

This document expands the current set of IBM System p5 documentation by providing a desktop reference that offers a detailed technical description of the p5-505.

This publication does not replace the latest IBM System p5 marketing materials, tools, or product documentation. It is intended as an additional source of information that, together with existing sources, can be used to enhance your knowledge of IBM server solutions.

The team that wrote this Redpaper

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

Charlie Cler is a Certified IT Specialist for IBM and has over 21 years of experience with IBM. He currently works in the United States as a pre-sales Systems Architect representing IBM Systems and Technology Group product offerings. He has been working with IBM System p servers for over 16 years.

Carlo Costantini is a Certified IT Specialist for IBM and has over 28 years of experience with IBM and IBM Business Partners. He currently works in Italy Pre-sales Field Technical Sales Support for IBM Sales Representatives and IBM Business Partners for all pSeries® and IBM eServer™ p5 systems offerings. He has broad marketing experience. He is a certified specialist for pSeries and IBM System p™ servers.

Michal Wawrzynski is a Sales Support Specialist in Poland. He has six years of experience in the IT industry. He has worked at IBM for two years selling and providing pre-sales support for IBM eServer p5, pSeries, OpenPower™, and IBM TotalStorage® systems. He has written extensively about system architecture and virtualization.

Yan ZHANG is an Advisory IT Specialist in IBM China. She has 14 years of experience in the IT field. She has worked at IBM for 11 years. Her areas of expertise include selling and providing pre-sales support for IBM System p5, eServer p5, pSeries, OpenPower, RS/6000®, and IBM TotalStorage systems. She has written extensively about system architecture, services, and reliability, availability, and serviceability (RAS).

The project that produced this document was managed by:
Scott Vetter

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

Christopher J. Algozzine, Stephen Hall, John Hilburn, Lindy Legler, Bill Mihaltse, Thoi Nguyen, Jan Palmer, Philip W. Sobey, Mike Stys, and Doug Szerdi
IBM U.S.

Gregor Linzmeier and Volker Haug
IBM Germany

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 or clients.

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

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

ibm.com/redbooks/residencies.html

Comments welcome

Your comments are important to us!

We want our papers to be as helpful as possible. Send us your comments about this Redpaper 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. HYTD Mail Station P099
  2455 South Road
  Poughkeepsie, NY 12601-5400

1

General description

The IBM System p5 505 and p5 505Q rack-mount server (9115-505) supports applications such as file-and-print, Web serving, networking, systems management, and security. Its IBM POWER5+™ processor is also ideally suited for high-performance compute clusters. To simplify naming, this product is referred to as the p5-505.

The p5-505 server comes in a 1U rack drawer package and is available in a one-core or two-core configuration using 64-bit, copper-based, and Silicon-On-Insulator (SOI) IBM POWER5+ microprocessors.

The p5-505 offers several POWER5+ processor options with the 1-core 1.9 GHz processor with no L3 cache, the 2-core 1.9 GHz processor with a 36 MB L3 cache, and the 2-core 2.1 GHz processor with a 36 MB L3 cache. The System p5 505Q also offers a processor option with the 4-core 1.65 GHz processor with two 36 MB L3 caches. The processors on the p5-505 allow you to configure either a 1-core or 2-core system, and the processor on the p5-505Q allows you to configure a 4-core system.

The p5-505 has a base of 1 GB of main memory that can be expanded to 32 GB. The p5-505 contains three internal device bays. These three device bays are front-accessible; two bays are for hot-swap-capable 3.5-inch disk drives and can accommodate up to 600 GB of disk storage. The third bay is available for a slim-line DVD-ROM or DVD-RAM. Other integrated features include two 64-bit PCI-X slots, an integrated service processor, integrated 10/100/1000 Mbps two-port Ethernet, one system port, two USB and two HMC ports, integrated dual channel Ultra320 SCSI controller, external SCSI port, hot-swappable power and cooling (redundant), and optional redundant power.

For partitioning, a Hardware Management Console (HMC) is recommended, but it is not required with the Integrated Virtualization Manager available with IBM Virtual I/O Server Version 1.3. Dynamic LPAR is supported on the p5-505, allowing up to two logical partitions using dedicated resources. In addition, the optional Advanced POWER™ Virtualization hardware feature supports up to 20 micropartitions using Micro-Partitioning™ technology.

IBM Cluster Systems Management V1.5.1 for AIX 5L and Linux and High-Availability Cluster Multi-Processing (HACMP™) for AIX 5L are supported on the p5-505.

The p5-505 is backed by a three-year limited warranty. Check with your IBM representative for the particular warranty availability in your region.

1.1 System specifications

Table 1-1 lists the general system specifications of the p5-505.

Table 1-1 IBM System p5 505 server specifications

Description	Range
Operating temperature	5 to 35 degrees Celsius (41 to 95 degrees Farenheit)
Relative humidity	8% to 80%
Operating voltage	100-127 (12 A) or 200-240 volts ac (20A) auto-ranging
Operating frequency	50/60 plus or minus 0.5 Hz
Maximum power consumption	500 watts
Maximum thermal output	1707 BTU/hr (British Thermal Unit)

1.2 Physical package

Table 1-2 lists the major physical attributes found on the p5-505.

Table 1-2 IBM System p5 505 server physical packaging

Dimension	
Height	43 mm (1.7 inches)
Width	440 mm (17.3 inches)
Depth	710 mm (28.0 inches)
Weight	
Minimum configuration	17 kg (37 lb)
Maximum configuration	23.2 kg (51 lb)

1.3 Rack-mount model

The p5-505 is available only as a rack-mount model.

Figure 1-1 shows a p5-505 that has been removed from a rack.

Figure 1-1 The p5-505 rack-mount server removed from the rack

The p5-505 is a 1U high, rack-mount server, designed to be installed in a 19-inch rack. There is no deskside model available. One of the following feature codes (FCs) must be ordered with the system:

- FC 7927 IBM Rack-mount Drawer Bezel and Hardware
- FC 7932 OEM Rack-mount Drawer Bezel and Hardware

The p5-505 can be installed in either IBM or OEM racks. For OEM rack requirements, see 1.6, “System racks” on page 8. There is one adjustable rack-mount drawer rail kit available for both IBM and OEM racks: FC 7103 IBM/OEM Rack-mount Drawer Rail Kit.

To ease cable management, there are 6 ft, 9 ft, and 14 ft jumper power cords (between the Central Electronic Complex, or CEC, drawer and the power distribution unit, or PDU) available and a set of cable-management arms.

Included with the p5-505 rack-mount packaging are all of the components and instructions necessary to enable installation in a 19-inch rack.

Figure 1-2 shows a more detailed view of the p5-505 rack-mount server, including connectors, location codes, SCSI IDs, and major components.

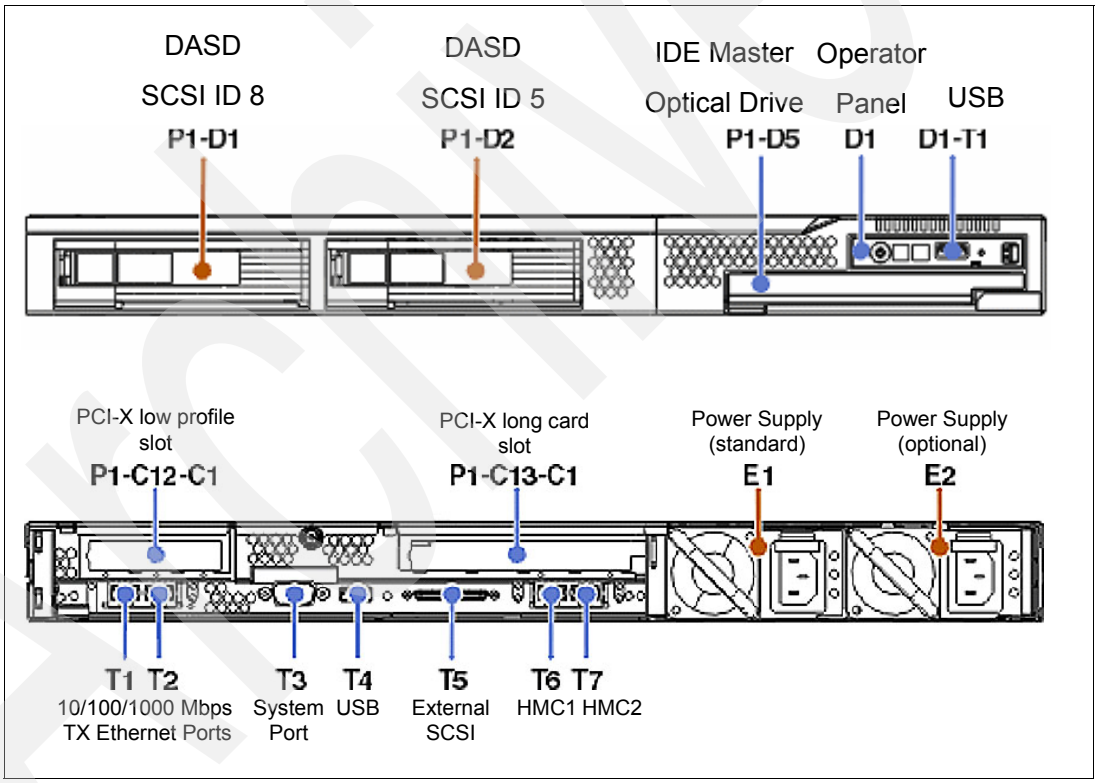


Figure 1-2 Front and rear view of the p5-505 including location codes

1.4 Minimum and optional features

The IBM System p5 505 Express Product Offering is based on a flexible, modular design featuring:

- 1-core, 2-core symmetric multiprocessor (SMP) design using one POWER5+ or POWER5™ processor and 4-core symmetric multiprocessor (SMP) design using one POWER5+ packaged in a processor module soldered directly to the system planar
- 1 GB of 533 MHz DDR-2 error checking and correcting (ECC) memory, expandable to 32 GB
- Two hot-swappable disk drive bays
- Two 64-bit, 3.3 volt, 266 MHz PCI-X 2.0 slots
- One slim-line media bay
- Service processor
- Redundant and hot-swap power supplies
- Hot-plug and redundant fans

The p5-505 includes the following integrated ports:

- Dual port 10/100/1000 Ethernet
- Integrated Dual Channel Ultra320 SCSI controller (one internal and one external VHDCI LVD connector)
- Two USB ports
- One system port
- Two HMC ports

The system supports 32-bit and 64-bit applications and requires specific levels of AIX 5L and Linux operating systems. See 2.13, “Operating system support” on page 50.

1.4.1 Processor features

The p5-505 and p5-505Q servers feature one or two POWER5+ processors with one, two, or four active processor cores running at 2.1 or 1.9 GHz (1-core and 2-core) or 1.65 GHz (4-core). Processors are installed on either a single-core module (SCM), dual-core module (DCM), or quad-core module (QCM). For a list of available processor features, refer to Table 1-3.

Table 1-3 Available processor options

Feature code	Description
8289	1-core 1.9 GHz POWER5+ Processor Card, no L3 cache
7679	2-core 1.9 GHz POWER5+ Processor Card, 36 MB L3 cache
8290	2-core 2.1 GHz POWER5+ Processor Card, 36 MB L3 cache
8288	4-core 1.65 GHz POWER5+ Processor Card, two 36 MB L3 caches
7650	1-core 1.65 GHz POWER5 Processor Card, no L3 cache
7652	2-core 1.65 GHz POWER5 Processor Card, 36 MB L3 cache
7674	2-core 1.5 GHz POWER5 Processor Card, 36 MB L3 cache

1.4.2 Memory features

The minimum memory requirement for the p5-505 is 1 GB, and the maximum capacity is 32 GB using 533 MHz DDR-2 DIMMs that are operating at 528 MHz. Memory DIMMs are installed into eight DIMM sockets that are located on the system planar. DIMMs can be installed in pairs or quad DIMM configurations. Note that an amount of memory is always in use by the POWER Hypervisor™, even when the machine is not partitioned. The System Planning Tool can be used to calculate the amount of available memory for an operating system based on machine configuration:

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

Table 1-4 lists the available memory features.

Table 1-4 Memory feature codes

Feature code	Description
1930	1 GB (2 x 512 MB) DIMMs, 276-pin DDR-2, 533 MHz SDRAM
1931	2 GB (2 x 1 GB) DIMMs, 276-pin DDR-2, 533 MHz SDRAM
1932	4 GB (2 x 2 GB) DIMMs, 276-pin DDR-2, 533 MHz SDRAM
1934	8 GB (2 x 4 GB) DIMMs, 276-pin DDR-2, 533 MHz SDRAM

1.4.3 Disk and media features

The p5-505 features two disk bays and one slim-line media bay. The minimum configuration requires at least one disk drive. Table 1-5 shows the disk drive feature codes that each bay can contain.

Table 1-5 Hot-swappable disk drive options

Feature code	Description
1968	73.4 GB ULTRA320 10 K rpm SCSI hot-swappable disk drive
1969	146.8 GB ULTRA320 10 K rpm SCSI hot-swappable disk drive
1970	36.4 GB ULTRA320 15 K rpm SCSI hot-swappable disk drive
1971	73.4 GB ULTRA320 15 K rpm SCSI hot-swappable disk drive
1972	146.8 GB ULTRA320 15 K rpm SCSI hot-swappable disk drive
1973	300 GB ULTRA320 10 K rpm SCSI hot-swappable disk drive

Either the DVD-ROM or DVD-RAM drive can be installed in the slim-line bay:

- FC 1903, DVD-ROM drive
- FC 1900, DVD-RAM drive

A logical partition that is running a supported release of the Linux operating system requires a DVD drive to provide a method for running the hardware diagnostic from the CD. Concurrent diagnostics, as provided by the AIX 5L **diag** command, are not available in the Linux operating system at the time of writing.

An internal redundant array of independent disks (RAID) enablement feature, FC 1908, is also available.

1.4.4 USB diskette drive

The externally attached USB diskette drive provides storage capacity up to 1.44 MB (FC 2591) on high-density (2HD) floppy disks and 720 KB on a double density floppy disk. It includes a 350 mm (13.7 in.) cable with standard USB connector. This super slim-line and lightweight USB V2-attached diskette drive takes its power requirements from the USB port. The drive can be attached to the integrated USB ports, or to a USB adapter (FC 2738). A maximum of one USB diskette drive is supported per integrated controller/adapter. The same controller can share a USB mouse and keyboard.

1.4.5 Hardware Management Console models

A p5-505 and p5-505Q server can be either HMC-managed or non-HMC-managed. In HMC-managed mode, an HMC is required as a dedicated workstation that allows you to configure and manage partitions. The HMC provides a set of functions to manage the system LPARs, dynamic LPAR operations, virtual features, Capacity on Demand, inventory and microcode management, and remote power control functions. These functions also include the handling of the partition profiles that define the processor, memory, and I/O resources that are allocated to an individual partition.

Note: Non-HMC-managed modes:

- Are full system partition mode (only one partition with all system resources).
- Use the Integrated Virtualization Manager (see 2.11.5, “Integrated Virtualization Manager” on page 44).

See 2.12, “Hardware Management Console” on page 47 for detailed information about the HMC.

Table 1-6 lists the HMC options for POWER5+ processor-based systems that are available at the time of writing. Existing HMC models can be also used.

Table 1-6 Supported HMC options

Type-model	Description
7310-C05	IBM 7310 Model C05 Deskside Hardware Management Console
7310-CR3	IBM 7310 Model CR3 Rack-Mount Hardware Management Console

Systems require Ethernet connectivity between the HMC and one of the Ethernet ports of the service processor. Ensure that sufficient HMC Ethernet ports are available to enable public and private networks if you need both. The 7310 Model C05 is a deskside model, which has one native 10/100/1000 Ethernet port. They can be extended with two additional two-port 10/100/1000 Gb adapters. The 7310 Model CR3 is a 1U, 19-inch rack mountable drawer that has two native Ethernet ports and can be extended with one additional two-port 10/100/1000 Gb adapter.

When an HMC is connected to the server, the integrated system ports are disabled. If you need serial connections, for example, non-Ethernet HACMP heartbeat, you must provide an Async adapter (FC 5723 or FC 2943).

Note: It is not possible to connect POWER4™ systems with POWER5, and POWER5+ processor-based systems simultaneously to the same HMC, but it is possible to connect POWER5 and POWER5+ processor-based systems together to the same HMC.

1.5 Express Product Offerings

The Express Product Offerings are a convenient way to order any of several configurations that are designed to meet typical client requirements. Special reduced pricing is available when a system order satisfies specific configuration requirements for memory, disk drives, and processors.

1.5.1 Express Product Offering requirements

When you order an Express Product Offering, the configurator offers a choice of starting points that can be added to. Clients can configure systems with 1-core, 2-core, or 4-core processors and up to 4 processor activations.

With the purchase of an Express Product Offering, for each paid processor activation, the client is entitled to one processor activation at no additional charge, if the following requirements are met:

- The system must have at least two disk drives of at least 73.4 GB each.
- There must be at least 1 GB of memory installed for each active processor.

When you purchase an Express Product Offering, you are entitled to a lower priced AIX 5L or Linux operating system license, or you can choose to purchase the system with no operating system. The lower priced AIX 5L or Linux operating system is processed through a feature number on AIX 5L and either Red Hat or SUSE Linux operating system. You can choose either the lower priced AIX 5L or Linux subscription, but not both.

If you choose AIX 5L for your lower priced operating system, you can also order Linux but you must purchase your Linux subscription at full price versus the reduced price. The same is true if you choose a Linux subscription as your lower priced operating system. Systems with a reduced price AIX 5L offering are the IBM System p5 Express Product Offering, AIX 5L edition, and systems with a lower priced Linux operating system are referred to as the IBM System p5 Express Product Offering, OpenPower edition.

In the case of Linux, only the first subscription purchased is lower priced so, for example, additional licenses purchased for Red Hat to run in multiple partitions are full price.

You can make changes to the standard features as needed and still qualify for processor entitlements at no additional charge and a reduced price AIX 5L or Linux operating system license. However, selection of total memory or DASD smaller than the total defined as the minimum disqualifies the order as an Express Product Offering.

1.5.2 Configurator starting points for Express Product Offerings

All product offerings have a set of standard features for the rack-mounted and deskside versions as listed in Table 1-7 on page 8 through Table 1-9 on page 8.

Table 1-7 Express Product Offering standard set of feature codes

Feature code description	Rack-mounted feature codes
Rack-mount bezel and hardware	7927 x 1
600 Watt power supply	7958 x 1
IDE DVD-ROM	1903 x 1
73.4 GB 10 k disk drives	1968 x 2
Language group specify	9300 or 93XX
Power cord	Select correct feature code

Table 1-8 POWER5+ Express Product Offering features - SCM, DCM, and QCM configurations

Description	1.9 GHz		2.1 GHz	1.65 GHz
Configuration	1-core	2-core	2-core	4-core
Processor cards	8289 x 1	7679 x 1	8290 x 1	8288 x 1
Processor activations	n/a	7689 x 1	7287 x 1	7288 x 2
Zero-priced Express Product Offering activations	8489 x 1	8487 x 1	8490 x 1	8488x 2
Total active processors	1	2	2	4
Minimum memory	1 GB	2 GB	2 GB	4 GB

Table 1-9 POWER5 Express Product Offering features - SCM and DCM configurations

Description	1.5 GHz	1.65 GHz	
Configuration	2-core	1-core	2-core
Processor cards	7674 x 1	7650 x 1	7652 x 1
Processor activations	7574 x 1	n/a	7372 x 1
Zero-priced Express Product Offering activations	8634 x 1	8639 x 1	8641 x 1
Total active processors	2	1	2
Minimum memory	2 GB	1 GB	2 GB

1.6 System racks

The IBM 7014 Model S11, S25, T00, and T42 Racks are 19-inch racks for general use with IBM System p rack-mount servers. The racks provide increased capacity, greater flexibility, and improved floor space utilization.

If a System p5 server is to be installed in a non-IBM rack or cabinet, you must ensure that the rack conforms to the EIA[1] standard EIA-310-D (see 1.6.9, “OEM rack” on page 17).

[1] Electronic Industries Alliance (EIA). Accredited by American National Standards Institute (ANSI), EIA provides a forum for industry to develop standards and publications throughout the electronics and high-tech industries.

Note: It is the client's responsibility to ensure that the installation of the drawer in the preferred rack or cabinet results in a configuration that is stable, serviceable, safe, and compatible with the drawer requirements for power, cooling, cable management, weight, and rail security.

1.6.1 IBM 7014 Model T00 Rack

The 1.8-meter (71-inch) Model T00 is compatible with past and present IBM System p servers. It is a 19-inch rack and is designed for use in all situations that have previously used the earlier rack models R00 and S00. The T00 rack has the following features:

- 36 EIA units (36U) of usable space.
- Optional removable side panels.
- Optional highly perforated front door.
- Optional side-to-side mounting hardware for joining multiple racks.
- Standard business black or optional white color in OEM format.
- Increased power distribution and weight capacity.
- Optional reinforced (ruggedized) rack feature (FC 6080) provides added earthquake protection with modular rear brace, concrete floor bolt-down hardware, and bolt-in steel front filler panels.
- Support for both ac and dc configurations.
- The dc rack height is increased to 1926 mm (75.8 inches) if a power distribution panel is fixed to the top of the rack.
- Up to four power distribution units (PDUs) can be mounted in the PDU bays (see Figure 1-3 on page 13), but others can fit inside the rack. See 1.6.6, "The ac power distribution unit and rack content" on page 12.
- Weights:
  - T00 base empty rack: 244 kg (535 pounds)
  - T00 full rack: 816 kg (1795 pounds)

1.6.2 IBM 7014 Model T42 Rack

The 2.0-meter (79.3-inch) Model T42 addresses the client requirement for a tall enclosure to house the maximum amount of equipment in the smallest possible floor space. The features that differ in the Model T42 rack from the Model T00 include:

- 42 EIA units (42U) of usable space (6U of additional space).
- The Model T42 supports ac only.
- Weights:
  - T42 base empty rack: 261 kg (575 pounds)
  - T42 full rack: 930 kg (2045 pounds)

Optional Rear Door Heat eXchanger (FC 6858)

Improved cooling from the heat exchanger enables the client to populate individual racks more densely, freeing valuable floor space without the need to purchase additional air conditioning units. The Rear Door Heat eXchanger features:

- A water-cooled heat exchanger door that is designed to dissipate heat generated from the back of computer systems before it enters the room

- An easy-to-mount rear door design that attaches to client-supplied water, using industry standard fittings and couplings
- Up to 15 KW (approximately 50,000 BTUs/hr) of heat removed from air exiting the back of a fully populated rack
- One year, limited warranty

Physical specifications

The physical specifications for the Rear Door Heat eXchanger are:

- Approximate height: 1.945.5 mm (76.6 inches)
- Approximate width: 635.8 mm (25.03 inches)
- Approximate depth: 141.0 mm (5.55 inches)
- Approximate weight: 31.9 kg (70.0 lb)

Client responsibilities

The client responsibilities are:

- Secondary water loop (to building chilled water)
- Pump solution (for secondary loop)
- Delivery solution (hoses and piping)
- Connections: standard 3/4 inch internal threads

1.6.3 IBM 7014 Model S11 Rack

The Model S11 rack can satisfy many light-duty requirements for organizing smaller rack-mount servers and expansion drawers. The 0.6-meter-high rack has:

- A perforated, lockable front door
- A heavy-duty caster set for easy mobility
- A complete set of blank filler panels for a finished look
- EIA unit markings on each corner to aid assembly
- A retractable stabilizer foot

The Model S11 rack has the following specifications:

- Width: 520 mm (20.5 inches) with side panels
- Depth: 874 mm (34.4 inches) with front door
- Height: 612 mm (24.0 inches)
- Weight: 37 kg (75.0 lb)

The S11 rack has a maximum load limit of 16.5 kg (36.3 lb) per EIA unit for a maximum loaded rack weight of 216 kg (475 lb).

1.6.4 IBM 7014 Model S25 Rack

The 1.3-meter-high Model S25 Rack will satisfy many light-duty requirements for organizing smaller rack-mount servers. Front and rear rack doors include locks and keys, helping keep your servers secure. Side panels are a standard feature, simplifying ordering and shipping. This 25U rack can be shipped configured and can accept server and expansion units up to 28-inches deep.

The front door is reversible and can be configured for either left or right opening. The rear door is split vertically in the middle and hinges on both the left and right sides. The S25 rack has the following specifications:

- Width: 605 mm (23.8 in.) with side panels
- Depth: 1001 mm (39.4 in.) with front door
- Height: 1344 mm (49.0 in.)
- Weight: 100.2 kg (221.0 lb)

The S25 rack has a maximum load limit of 22.7 kg (50 lb) per EIA unit for a maximum loaded rack weight of 667 kg (1470 lb).

1.6.5 S11 rack and S25 rack considerations

The S11 and S25 racks do not have vertical mounting space that will accommodate feature number 7188 PDUs. All PDUs required for application in these racks must be installed horizontally in the rear of the rack. Each horizontally mounted PDU occupies 1U of space in the rack and therefore reduces the space available for mounting servers and other components.

FC 0469 Client Specified Rack Placement provides the client the ability to specify the physical location of the system modules and attached expansion modules (drawers) in the racks. The client's input is collected and verified through the marketing configurator (eConfig). The client's request is reviewed by eConfig for safe handling by checking the weight distribution within the rack. The Manufacturing Plant provides the final approval for the configuration. This information is then used by IBM Manufacturing to assemble the system components (drawers) in the rack according to the client's request.

The CFReport from eConfig must be submitted to the following site:

`http://www.ibm.com/servers/eserver/power/csp`

Table 1-10 lists the machine types supported in the S11 and S25 racks.

Table 1-10 Models supported in S11 and S25 racks

Machine type-model	Name	Supported in	
		7014-S11 rack	7014-S25 rack
7037-A50	System p5 185	X	X
7031-D24/T24	EXP24 Disk Enclosure	X	X
7311-D20	I/O Expansion Drawer	X	X
9110-510	System p5 510	X	X
9111-520	System p5 520	X	X
9113-550	System p5 550	X	X
9115-505	System p5 505	X	X
9123-710	OpenPower 710	X	X
9124-720	OpenPower 720	X	X
9110-510	System p5 510 and 510Q	X	X
9131-52A	System p5 520 and 520Q	X	X

Machine type-model	Name	Supported in	
		7014-S11 rack	7014-S25 rack
9133-55A	System p5 550 and 550Q	X	X
9116-561	System p5 560Q	X	X
9910-P33	3000 VA UPS (2700 watt)	X	X
9910-P65	500 VA UPS (208-240V)		X
7315-CR3	Rack-mount HMC		X
7310-CR3	Rack-mount HMC		X
7026-P16	LAN-attached remote asynchronous node (RAN)		X
7316-TF3	Rack-mounted flat-panel console kit		X

1.6.6 The ac power distribution unit and rack content

Note: Each server, or a system drawer to be mounted in the rack, requires two power cords that are not included in the base order. For maximum availability, we highly recommend to connect power cords from the same server or system drawer to two separate PDUs in the rack. These PDUs could be connected to two independent client power sources.

For rack models T00 and T42, 12-outlet PDUs (FC 9188 and FC 7188) are available. For rack models S11 and S25, FC 7188 is available.

Four PDUs can be mounted vertically in the T00 and T42 racks. Figure 1-3 on page 13 shows the placement of the four vertically mounted PDUs. In the rear of the rack, two additional PDUs can be installed horizontally in the T00 rack and three in the T42 rack. The four vertical mounting locations are filled first in the T00 and T42 racks. Mounting PDUs horizontally consumes 1U per PDU and reduces the space available for other racked components. When mounting PDUs horizontally, we recommend that you use fillers in the EIA units that are occupied by these PDUs to facilitate proper air-flow and ventilation in the rack.

The S11 and S25 racks support as many PDUs as there is available rack space.

For detailed power cord requirements and power cord feature codes, see *IBM System p5, @server p5 and i5, and OpenPower Planning*, SA38-0508. For an online copy, select **Map of pSeries books to the information center → Planning → Printable PDFs → Planning** at the following Web site:

http://publib.boulder.ibm.com/infocenter/eserver/v1r3s/index.jsp

Note: Ensure that the appropriate power cord feature is configured to support the power that is being supplied.

The Base/Side Mount Universal PDU (FC 9188) and the optional, additional Universal PDU (FC 7188) support a wide range of country requirements and electrical power specifications. The PDU receives power through a UTG0247 power line connector. Each PDU requires one PDU-to-wall power cord. Nine power cord features are available for different countries and applications by varying the PDU-to-wall power cord, which must be ordered separately. Each power cord provides the unique design characteristics for the specific power requirements.

To match new power requirements and preserve previous investments, these power cords can be requested with an initial order of the rack or with a later upgrade of the rack features.

The PDU has 12 client-usable IEC 320-C13 outlets. There are six groups of two outlets fed by six circuit breakers. Each outlet is rated up to 10 amps, but each group of two outlets is fed from one 15 amp circuit breaker.

Note: Based on the power cord that is used, the PDU can supply from 4.8 kVA to 19.2 kVA. The total kilovolt ampere (kVA) of all the drawers plugged into the PDU must not exceed the power cord limitation.

The Universal PDUs are compatible with previous IBM System p5, OpenPower, and pSeries models.

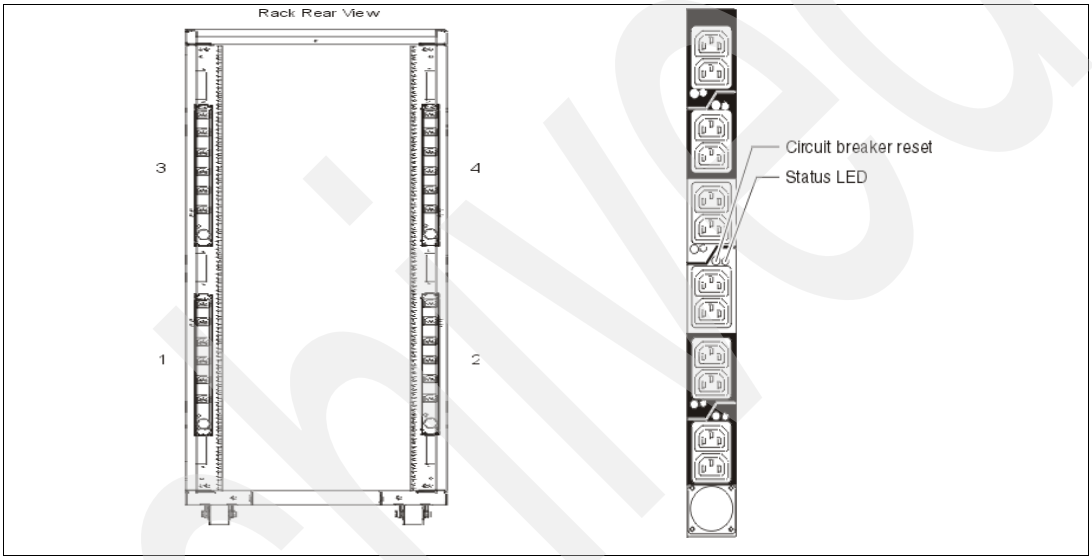


Figure 1-3 PDU placement and PDU view

1.6.7 Rack-mounting rules for the p5-505

The p5-505 is a 1U rack-mounted server drawer. Consider the following primary rules when mounting the p5-505 into a rack:

- The p5-505 is designed to be placed at any location in the rack. For rack stability, it is advisable to start filling a rack from the bottom.
- Any remaining space in the rack can be used to install other systems or peripherals, provided that the maximum permissible weight of the rack is not exceeded and the installation rules for these devices are followed.
- Before placing or sliding a p5-505 into the service position, it is essential that the rack manufacturer's safety instructions are followed regarding rack stability.
- Considering only the maximum configuration of a single p5-505, a maximum of 24 model p5-505 servers fit in the T00 rack, 28 p5-505 servers in the T42 rack, seven p5-505 servers in the S11 rack, and 24 p5-505 servers in the S25 rack due to weight and power requirements.

Note: Carefully consider the following specifications during your rack planning for all the drawers including servers, storage, and their peripherals to be installed in the system.

The T00 rack's maximum weight of drawers is 572 kg (1260 lb) and the T42 is 667 kg (1470 lb). The minimum weight of the p5-505 is 17.0kg (37 lb) and the maximum weight of the p5-505 is 23.2 kg (51 lb). Do not exceed the rack's maximum weight of drawers.

There is a cable-management arm shipped with the p5-505 (refer to Figure 2-17 on page 58), which helps you better arrange the cables at the back of the p5-505. There are power cables, which have different lengths from the drawer to the PDU, available (5 ft., 9 ft., and 14 ft.).

1.6.8 Additional options for rack

This section highlights solutions that are available to provide a single point of management for environments composed of multiple System p5-505 or p5-505Q servers or other IBM System p servers.

IBM 7212 Model 103 IBM TotalStorage storage device enclosure

The IBM 7212 Model 103 is designed to provide efficient and convenient storage expansion capabilities for selected System p servers. The IBM 7212 Model 103 is a 1U rack-mountable option to be installed in a standard 19-inch rack using an optional rack-mount hardware feature kit. The 7212 Model 103 has two bays that can accommodate any of the following storage drive features:

- Digital Data Storage (DDS) Gen 5 DAT72 Tape Drive provides a physical storage capacity of 36 GB (72 GB with 2:1 compression) per data cartridge.
- VXA-2 Tape Drive provides a media capacity of up to 80 GB (160 GB with 2:1 compression) physical data storage capacity per cartridge.
- VXA-320 Tape Drive provides a media physical capacity of up to 160 GB (320 GB with 2:1 compression) physical data storage capacity per cartridge.
- Half-High LTO-2 Tape Drive provides media physical capacity of up to 200 GB (400 GB with 2:1 compression) data storage per Ultrium 2 cartridge and a sustained data transfer rate of 24.0 MB per second (48 MB per second with 2:1 compression). In addition to reading and writing on Ultrium 2 tape cartridges, it is also read and write compatible with Ultrium 1 cartridges.
- SLR60 Tape Drive (QIC format) comes with a media with 37.5 GB native data physical capacity per tape cartridge and a native physical data transfer rate of up to 4 MB per second and uses 2:1 compression to achieve a single tape cartridge physical capacity of up to 75 GB of data.
- SLR100 Tape Drive (QIC format) comes with a media with 50 GB native data physical capacity per tape cartridge and a native physical data transfer rate of up to 5 MB per second and uses 2:1 compression to achieve single tape cartridge storage of up to 100 GB of data.
- DVD-RAM 2 drive can read and write on 4.7 GB and 9.4 GB DVD-RAM media. The DVD-RAM 2 uses only bare media, which reduces media costs, and is also read compatible with multi-session CD, CD-RW, and 2.6 GB and 5.2 GB DVD-RAM media. The 9.4 GB physical capacity of DVD-RAM allows storage of more data than on conventional CD-R media. Fast performance also allows quick access to information, while downward compatibility helps provide investment protection.

Note: Disc capacity options are 2.6 GB and 4.7 GB per side. The 5.2 GB and 9.4 GB capacities can be achieved by using double-sided DVD-RAM discs.

Flat panel display options

The IBM 7316-TF3 Flat Panel Console Kit can be installed in the system rack. This 1U console uses a 17-inch thin film transistor (TFT) LCD with a viewable area of 337.9 mm x 270.03 mm and a 1280 x 1024 pel[2] resolution. The 7316-TF3 Flat Panel Console Kit has the following attributes:

- A 17-inch, flat screen TFT color monitor that occupies 1U (1.75 inches) in a 19-inch standard rack.
- Ability to mount the IBM Travel Keyboard in the 7316-TF3 rack keyboard tray
- Support for the new 1x8 LCM switch (FC 4280), the Netbay LCM2 (FC 4279) with access to and control of as many as 64 servers and support of both USB and PS/2 server-side keyboard and mouse connections
- IBM Travel Keyboard mounts in the rack keyboard tray (Integrated Track point and UltraNav)

IBM PS/2 Travel Keyboards are supported on the 7316-TF3 for use in configurations where only PS/2 keyboard ports are available.

The IBM 7316-TF3 Flat Panel Console Kit provides an option for the USB Travel Keyboards with UltraNav. The keyboard enables the 7316-TF3 to be connected to systems that do not have PS/2 keyboard ports. The USB Travel Keyboard can be directly attached to an available integrated USB port or a supported USB adapter (2738) on System p5 servers or 7310-CR3 and 7315-CR3 HMCs.

The IBM 7316-TF3 flat-panel, rack-mounted console is now available with two console switch options, which lets you inexpensively cable, monitor, and manage your rack servers: the new 1x8 LCM Console Switch (FC 4280) and the LCM2 console switch (FC 4279).

The 1x8 Console Switch is a cost-effective, densely-packed solution that helps you set up and control selected System p rack-mounted IBM servers:

- Supports one local user with PS/2 keyboard, PS/2 mouse, and video connections
- Features an 8-port, CAT5 console switch for single-user local management
- Supports both USB and PS/2 server-side keyboard and mouse connections
- Occupies 1U (1.75 in) in a 19-inch standard rack

The 1x8 Console Switch can be mounted in one of the following racks: 7014-T00, 7014-T42, 7014-S11, or 7014-S25.

The 1x8 Console Switch supports GXT135P (FC 1980 and FC 2849) graphics accelerators. The following cables are used to attach the IBM servers to the 1x8 Console Switch:

- IBM 3M Console Switch Cable (PS/2) (FC 4282)
- IBM 3M Console Switch Cable (USB) (FC 4281)

The 1x8 Console Switch supports the following monitors:

- 7316-TF3 rack console monitor
- pSeries TFT monitors (FC 3641, FC 3643, FC 3644, and FC 3645)

[2] Picture elements

Separately available switch cables convert KVM signals for CAT5 cabling for servers with USB and PS/2 ports. A minimum of one cable feature (FC 4281) or USB feature (FC 4282) is required to connect the IBM 1x8 Console Switch (FC 4280) to a supported server. The 3-meter cable FC 4281 has one HD15 connector for video and one USB connector for keyboard and mouse. The 3-meter cable FC 4282 has one HD15 connector for video, one PS/2 connector for keyboard, and one PS/2 connector for mouse. It is used to connect the IBM 1x8 Console Switch to a supported server.

The 1x8 Console Switch is a 1U (1.75-inch) rack-mountable LCM switch containing eight analog rack interface ports for connecting switches using CAT5 cable. The switch supports a maximum video resolution of 1280x1024.

The Console Switch allows for two levels of tiering and supports up to 64 servers at a single user location through switch tiering. The previous VGA switch (FC 4200), the LCM switch (FC 4202), and LCM2 switch (FC 4279) can be tiered with the 1x8 Console Switch.

Note: When the 1x8 Console Switch is tiered with the previous VGA switch (FC 4200) or LCM (FC 4202) switch, it must be at the top level of the tier. When the 1x8 Console Switch is tiered with the LCM2 (FC 4279) switch, it must be at the secondary level of the tier.

The IBM Local 2x8 Console Manager (LCM2) switch (FC 4279) provides users single-point access and control of up to 1024 servers. It supports connection to servers with either PS/2 or USB connections with installation of appropriate options. The maximum resolution is 1280 x 1024 at 75Hz. The LCM2 switch can be tiered; three levels of tiering are supported.

A minimum of one LCM feature (FC 4268) or USB feature (FC 4269) is required with an IBM Local 2x8 Console Manager (LCM2) switch (FC 4279). Each feature can support up to four systems. When connecting to a p5-520 or p5-520Q, FC 4269 provides connection to the POWER5 USB ports. Only the PS/2 keyboard is supported when attaching the 7316-TF3 to the LCM Switch.

When selecting the LCM Switch, consider the following information:

- The KVM Conversion Option (KCO) cable (FC 4268) is used with systems with PS/2 style keyboard, display, and mouse ports.
- The USB cable (FC 4269) is used with systems with USB keyboard or mouse ports.
- The switch offers four ports for server connections. Each port in the switch can connect a maximum of 16 systems:
  - One KCO cable (FC 4268) or USB cable (FC 4269) is required for every four systems supported on the switch.
  - A maximum of 16 KCO cables or USB cables per port can be used with the Netbay LCM Switch to connect up to 64 servers.

Note: A server microcode update might be required on installed systems for boot-time System Management Services (SMS) menu support of the USB keyboards. For microcode updates, see:

`http://www14.software.ibm.com/webapp/set2/firmware/gjsn`

We recommend that you have the 7316-TF3 installed between EIA 20 and EIA 25 of the rack for ease of use. The 7316-TF3 or any other graphics monitor requires that a POWER GXT135P graphics accelerator (FC 1980 and FC 2849) is installed in the server, or some other graphics accelerator, if supported.

Hardware Management Console 7310 Model CR3

The 7310 Model CR3 Hardware Management Console (HMC) is a 1U, 19-inch rack-mountable drawer supported in the 7014 racks. For additional HMC specifications, see 2.12, “Hardware Management Console” on page 47.

1.6.9 OEM rack

The p5-505 can be installed in a suitable OEM rack, provided that the rack conforms to the EIA-310-D standard for 19-inch racks. This standard is published by the Electrical Industries Alliance, and a summary of this standard is available in the publication *IBM System p5, @server p5 and i5, and OpenPower Planning*, SA38-0508.

The key points mentioned in this documentation are as follows:

- The front rack opening must be 451 mm wide + 0.75 mm (17.73 inches + 0.03 inches), and the rail-mounting holes must be 465 mm + 0.8 mm (18.3 inches + 0.03 inches) apart on center (horizontal width between the vertical columns of holes on the two front-mounting flanges and on the two rear-mounting flanges). See Figure 1-4 for a top view showing the specification dimensions.

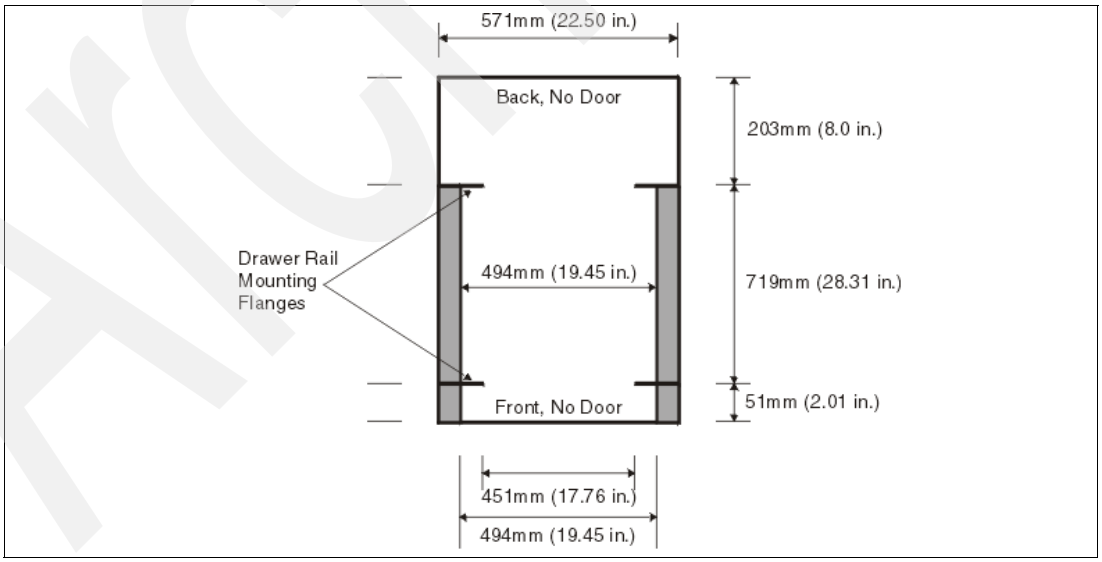


Figure 1-4 Top view of non-IBM rack specification dimensions

- The vertical distance between the mounting holes must consist of sets of three holes spaced (from bottom to top) 15.9 mm (0.625 inches), 15.9 mm (0.625 inches), and 12.67 mm (0.5 inches) on center, making each three-hole set of vertical hole spacing 44.45 mm (1.75 inches) apart on center. Rail-mounting holes must be 7.1 mm + 0.1 mm

(0.28 inches + 0.004 inches) in diameter. See Figure 1-5 on page 18 and Figure 1-6 on page 18 for the top and bottom front specification dimensions.

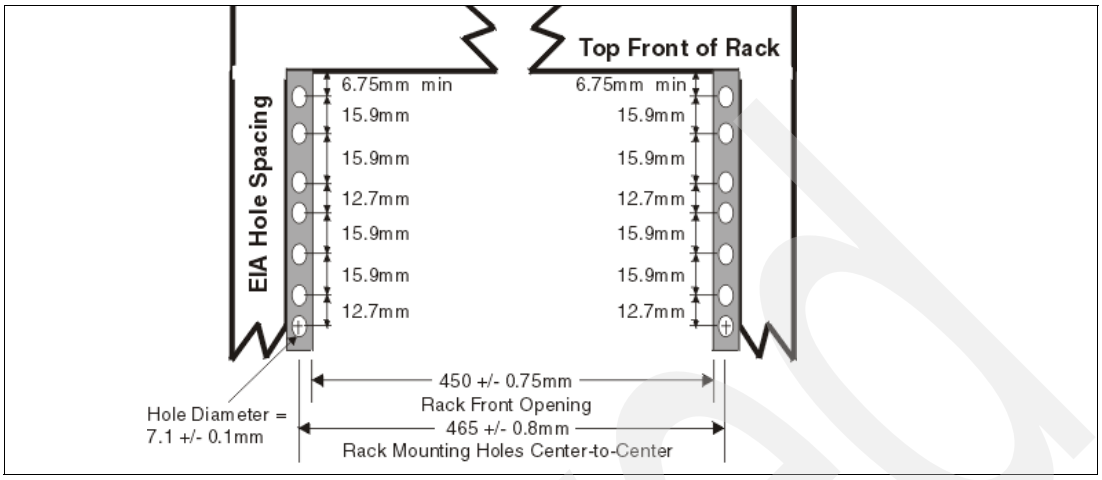


Figure 1-5 Rack specification dimensions, top front view

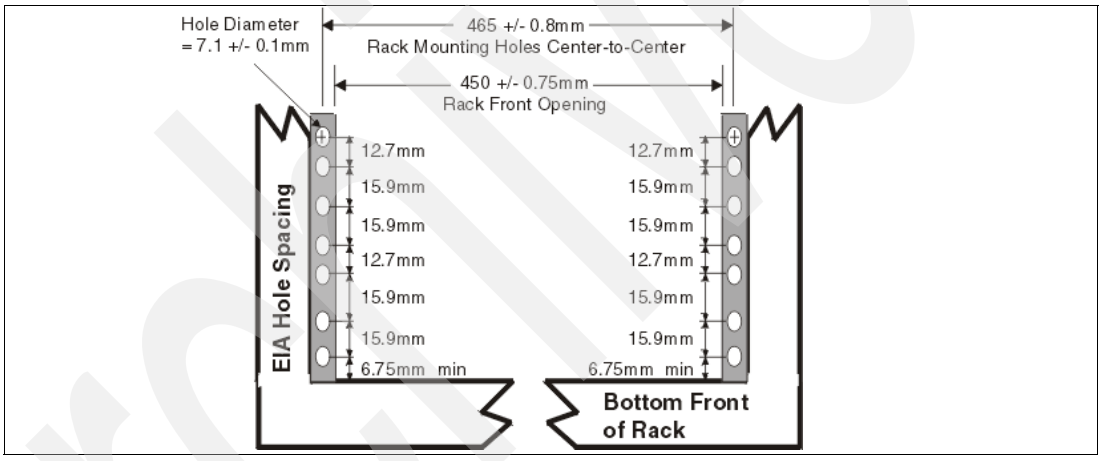


Figure 1-6 Rack specification dimensions, bottom front view

- It might be necessary to supply additional hardware, such as fasteners, for use in some manufacturer's racks.
- The system rack or cabinet must be capable of supporting an average load of 15.9 kg (35 lb) of product weight per EIA unit.
- The system rack or cabinet must be compatible with drawer mounting rails, including a secure and snug fit of the rail-mounting pins and screws into the rack or cabinet rail support hole.

Note: The OEM rack must only support ac-powered drawers. We strongly recommend that you use a power distribution unit (PDU) that meets the same specifications as the PDUs to supply rack power. Rack or cabinet power distribution devices must meet the drawer power requirements, as well as the requirements of any additional products that will be connected to the same power distribution device.

2

Architecture and technical overview

This chapter discusses the overall system architecture represented by Figure 2-1. This chapter describes the major components of this diagram in the following sections. The bandwidths provided throughout this section are theoretical maximums provided for reference. We recommend that you always obtain real-world performance measurements using production workloads.

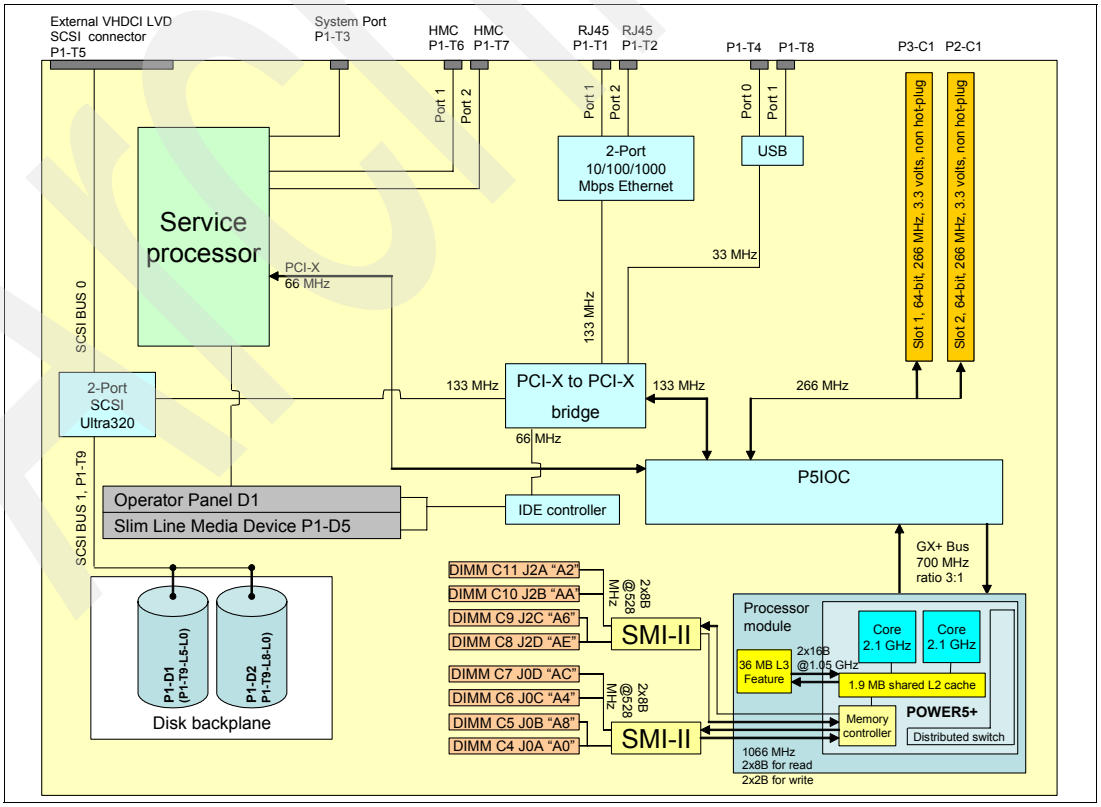


Figure 2-1 The p5-505 logic data flow with 2.1 GHz DCM

2.1 The POWER5+ processor

The IBM POWER5+ processor capitalizes on all the enhancements brought by the POWER5 processor. For a detailed description of the POWER5 processor, refer to *IBM System p5 505 Express Technical Overview and Introduction,* REDP-4079. Figure 2-2 shows a high level view of the POWER5+ processor.

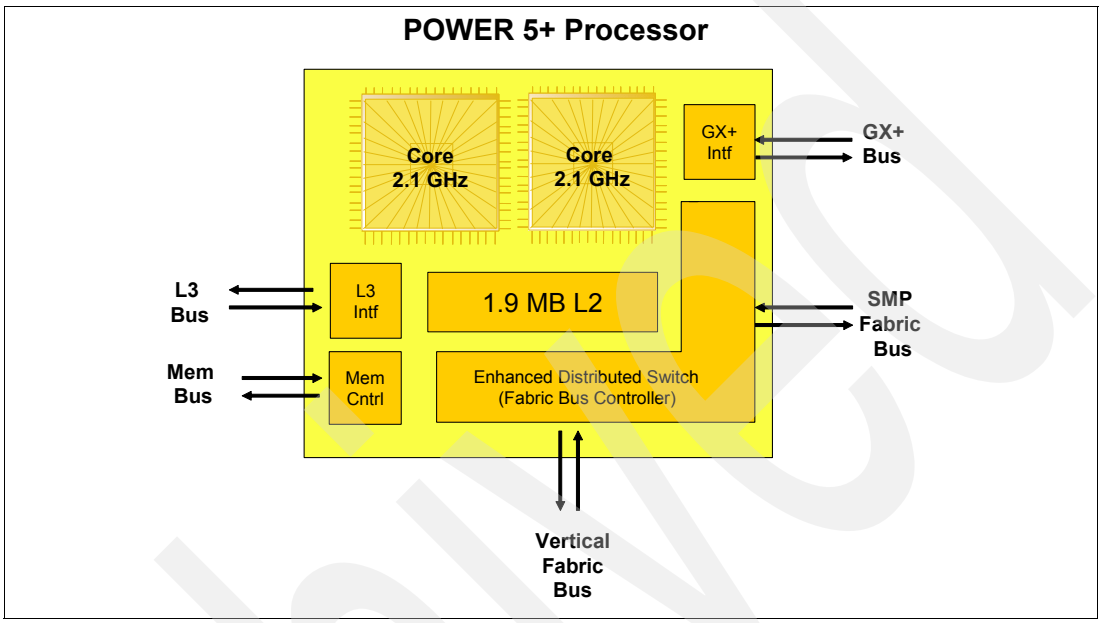


Figure 2-2 POWER5+ processor

The CMOS9S technology for the POWER5 processor used a 130 Nanometer (nm) fabrication process. The CMOS10S technology for the POWER5+ processor uses a 90 nm fabrication process, enabling:

- Performance gains through faster clock rates
- Physical size reduction (243 mm compared with 389 mm)

Compared to the POWER5 processor, the 37% smaller POWER5+ processor consumes less power and therefore requires less cooling. This allows it to be used in servers that previously used lower frequency processors because of cooling restrictions.

The POWER5+ design provides the following additional enhancements over its predecessor:

- New page sizes in ERAT and translation look-aside buffer (TLB) and two new page sizes (64 KB and 16 GB), which were recently added in PowerPC® architecture.
- New segment size in SLB and one new segment size (1 TB) that was recently added in PowerPC architecture.
- The doubling of the TLB size in the POWER5+ processor to 2048 entries.
- New floating-point round to integer instructions (frfin, frfiz, frfip, and frfim) that have been added to round floating-point numbers with the following rounding modes: nearest, zero, integer plus, and integer minus.
- Improved floating-point performance.
- Lock performance enhancement.
- Enhanced SLB read.

- True Little-Endian mode support as defined in the PowerPC architecture.
- Changes in the fabric, L2 and L3 controller, memory controller, GX controller, and RAS to provide support for the QCM that have resulted in SMP system sizes that are double what is available in POWER5 DCM-based servers. Current POWER5+ implementations support single address loop.
- Several enhancements have been made in the memory controller for improved performance. The memory controller is ready to support future DDR-2 667 MHz DIMMs.
- Enhanced redundancy in L1 Dcache, L2 cache, and L3 directory. Addition of:
  - Independent control of the L2 cache and the L3 directory for redundancy to allow split-repair action
  - Wordline redundancy in the L1 Dcache
  - Array Built-In Self Test (ABIST) column repair for the L2 cache and the L3 directory

2.2 Processor and cache

In the p5-505 and p5-505Q, the POWER5+ processors, associated L3 cache, and memory DIMMs are packaged on the system planar. The p5-505 and the p5-505Q use different POWER5+ processor modules.

Note: Because the POWER5+ and POWER5 processor modules are directly soldered to the system planar, special care must be taken in sizing and selecting the ideal CPU configuration.

2.2.1 POWER5+ single-core module

The 1-core p5-505 POWER5+ system planar contains a single-core module (SCM) and the local memory storage subsystem for that SCM. The POWER5+ single-core processor is packaged in the SCM. Figure 2-3 shows the layout of a p5-505 SCM and associated memory.

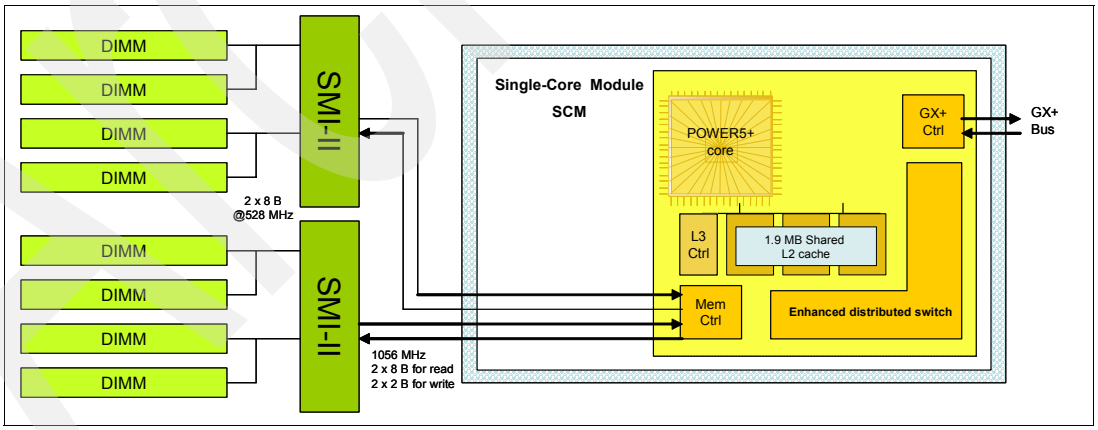


Figure 2-3 p5-505 POWER5+ 1.9 GHz SCM with DDR-2 memory socket layout view

The storage structure for the POWER5+ processor is a distributed memory architecture that provides high-memory bandwidth. The processor is interfaced to eight memory slots that are controlled by two Synchronous Memory Interface II (SMI-II) chips, which are located in close physical proximity to the processor module.

I/O connects to the p5-505 processor module using the GX+ bus. The processor module provides a single GX+ bus. The GX+ bus provides an interface to I/O devices through the RIO-2 connections.

2.2.2 POWER5+ dual-core module

The 2-core p5-505 POWER5+ system planar contains a dual-core module (DCM) and the local memory storage subsystem for that DCM. The POWER5+ dual-core processor and its associated L3 cache is packaged in the DCM.

Figure 2-4 shows a layout of a p5-505 DCM and associated memory.

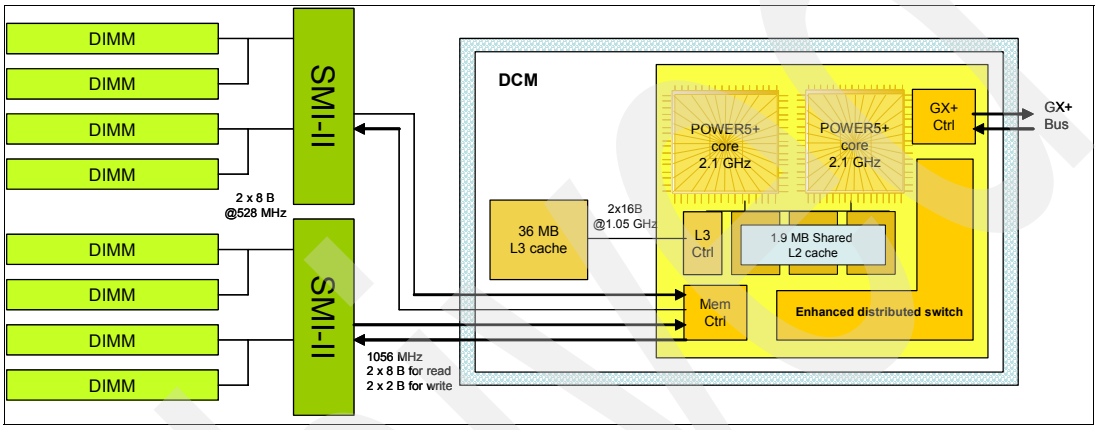


Figure 2-4 p5-505 POWER5+ 2.1 GHz DCM with DDR-2 memory socket layout view

The storage structure for the POWER5+ processor is a distributed memory architecture that provides high-memory bandwidth, although each processor can address all memory and sees a single shared memory resource. They are interfaced to eight memory slots that are controlled by two Synchronous Memory Interface II (SMI-II) chips, which are located in close physical proximity to the processor module.

I/O connects to the p5-505 processor using the GX+ bus. The processor provides a single GX+ bus. The GX+ bus provides an interface to I/O devices through the RIO-2 connections.

The theoretical maximum throughput of the L3 cache is 16-byte read, 16-byte write at a bus frequency of 1.05 GHz (based on a 2.1 GHz processor clock), which equates to 33600 MBps or 33.60 GBps. Further details are on Table 2-3 on page 26.

2.2.3 p5-505Q quad-core module

The 4-core p5-505Q system planar contains a quad-core module (QCM) and the local memory storage subsystem for that QCM. Two POWER5+ dual-core processors and their associated L3 Cache are packaged in the QCM. Figure 2-5 on page 23 shows a layout of p5-505Q QCM with associated memory.

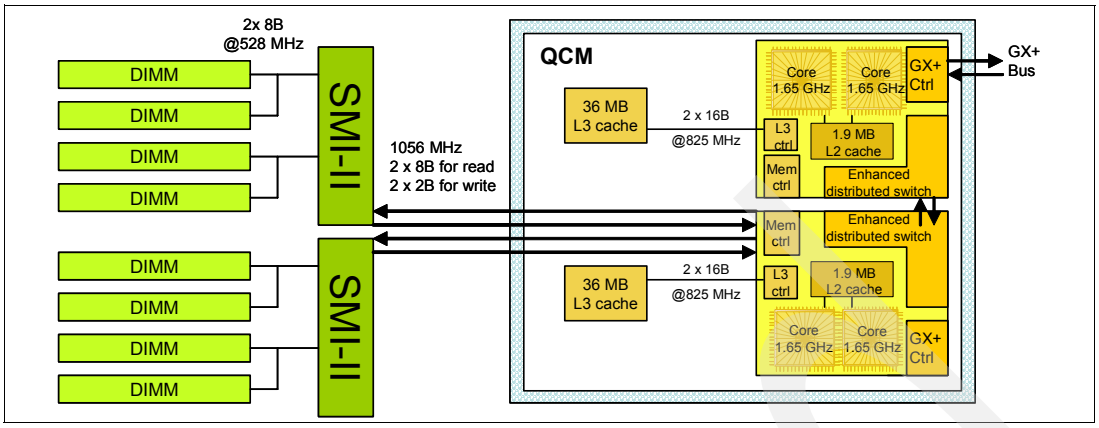


Figure 2-5 p5-505Q POWER5+ 1.65 GHz QCM with DDR-2 memory socket layout view

The storage structure for the POWER5+ processor is a distributed memory architecture that provides high-memory bandwidth. Each processor in the QCM can address all memory and sees a single shared memory resource. In the QCM, one POWER5+ processor has direct access to eight memory slots that are controlled by two SMI-II chips and located in close physical proximity to the processor module. The other POWER5+ processor has access to the same memory slots through the Vertical Fabric Bus.

I/O connects to the p5-505Q QCM using the GX+ bus. The QCM provides a single GX+ bus. Each processor in the POWER5+ processor has either a direct access to the GX+ Bus using its GX+ Bus controller or uses the Vertical Fabric Bus controlled by the Fabric Bus controller. The GX+ bus provides an interface to I/O devices through the RIO-2 connections. The POWER5+ processor that does not have direct access to memory does have a direct access to the GX+ Bus.

The theoretical maximum throughput of each L3 cache is 16-byte read, 16-byte write at a bus frequency of 825 MHz (based on a 1.65 GHz processor clock), which equates to 26400 MBps or 26.4 GBps per L3 cache. There are two L3 caches on the QCM, resulting in a total L3 cache theoretical maximum throughput of 52.8 GBps. Throughput rates are summarized in Table 2-3 on page 26.

2.2.4 Processor capacities and speeds

Table 2-1 describes the available processor capacities and speeds for the p5-505 and p5-505Q systems.

Table 2-1 The p5-505 and p5-505Q available processor capacities and speeds

	p5-505 @ 1.5 GHz	p5-505 @ 1.65 GHz	p5-505 @ 1.9 GHz	p5-505 @ 2.1 GHz	p5-505Q @ 1.65 GHz
1-core	No	Yes	Yes	No	No
2-core	Yes	Yes	Yes	Yes	No
4-core	No	No	No	No	Yes

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 is similar to the following output (False, as used in this output, signifies that the value cannot be changed through an AIX 5L command interface):

```
frequency 1498500000    Processor Speed       False
smt_enabled true        Processor SMT enabled False
smt_threads 2           Processor SMT threads False
state enable            Processor state       False
type powerPC_POWER5     Processor type        False
```

- **pmcycles -m**

  The **pmcycles** command (AIX 5L) uses the performance monitor cycle counter and the processor real-time clock to measure the actual processor clock speed in MHz. The following output is from a 2-core p5-505 system running at 1.5 GHz with simultaneous multithreading enabled:

```
Cpu 0 runs at 1498 MHz
Cpu 1 runs at 1498 MHz
```

Note: The **pmcycles** command is part of the bos.pmapi fileset. This component must be installed before using the **lslpp -l bos.pmapi** command.

2.3 Memory subsystem

The p5-505 and p5-505Q servers offer pluggable DDR-2 memory DIMMs. The rate of DDR-2 DIMMs is double that of DDR DIMMs (DDR DIMMs have double the rate bits of SDRAM), which enables up to four times the performance of traditional SDRAM. There are eight slots that are available on the system planar for up to eight pluggable DDR-2 DIMMs. The minimum memory for a server is 1.0 GB (2 x 512 MB) and 32 GB is the maximum installable memory. All memory is accessed by two of SMI-II chips that are located between memory and processor. The SMI-II supports multiple data flow modes.

2.3.1 Memory placement rules

In 1.4.2, “Memory features” on page 5, we list the memory features available at the time of writing for the p5-505 and p5-505Q.

Memory must be pluggable in pairs. All the memory features consist of two DIMMs. Memory feature numbers can be mixed within a system.

Table 2-2 on page 25 shows the memory installation rules. Memory must be balanced across the DIMM slots. The service information label, located on the top cover of the system, provides memory DIMMs slot location information.

Table 2-2 Memory installation rules in the p5-505

	Location order by slot	Preferred priority
Two-DIMM installation	C4, C11	1st
	C6, C9	2nd
	C5, C10	3rd
	C7, C8	4th
Four-DIMM installation	C4, C11, C6, C9	1st
	C5, C10, C7, C8	2nd
Six-DIMM installation	C4, C11, C6, C9, C5, C10	1st
	C6, C9, C7, C8, C4, C11	2nd

To determine how much memory is installed in a system, use the following command:

```
# lsattr -El sys0 | grep realmem
realmem     524288      Amount of usable physical memory in Kbytes False
```

Note: A quad must be made of identical DIMMs. Mixed DIMM capacities in a quad will result in reduced RAS.

2.3.2 OEM memory

OEM memory is not supported by IBM on the p5-505 or p5-505Q. OEM memory is not certified by IBM for use in System p servers. If the p5-505 or p5-505Q is populated with OEM memory, you might experience unexpected and unpredictable behavior, especially when the system is using Micro-Partitioning.

All IBM memory is identified by an IBM logo and a white label printed with a barcode and an alphanumeric string, illustrated in Figure 2-6.

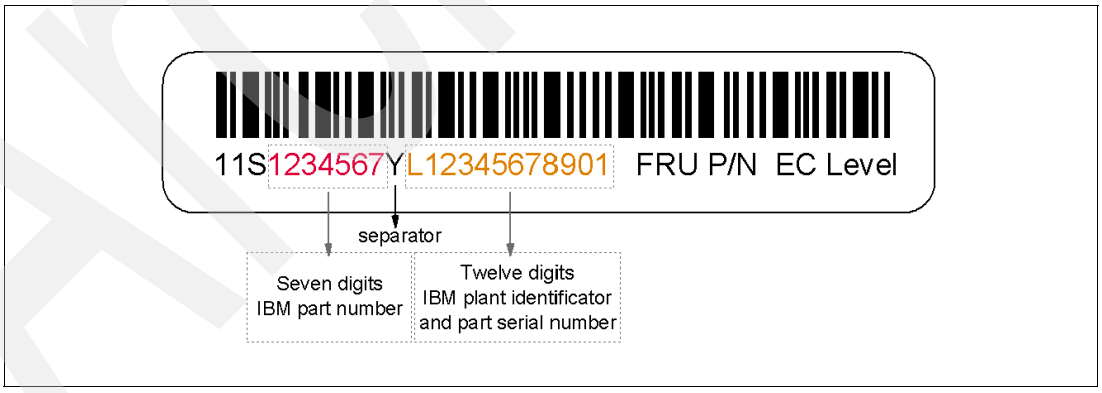


Figure 2-6 IBM memory certification label

2.3.3 Memory throughput

The memory subsystem throughput is based on the speed of the memory. An elastic interface in the POWER5+ processor buffers reads and writes to and from memory and the processor. There are two SMI-II components, each with a single 8-byte read and a 2-byte write, high- speed Elastic Interface-II bus to the memory controller of the processor. The bus allows double reads or writes per clock cycle. Because the bus operates at 1056 MHz, the

peak processor-to-memory throughput for read is (8 x 2 x 1056) = 16896 MBps or 16.89 GBps. The peak processor-to-memory throughput for write is (2 x 2 x 1056) = 4224 MBps or 4.22 GBps, for a total of 21.12 GBps.

The 533 MHz DDR-2 memory DIMMs operate at 528 MHz through four 8-byte paths. Read and write operations share these paths. There must be at least four DIMMs installed to effectively use each path. In this case, the throughput between the SMI-II and the DIMMs is (8 x 4 x 528) or 16.89 GBps.

These values are maximum theoretical throughputs for comparison purposes only. Table 2-3 provides the theoretical throughput values for different configurations.

Table 2-3 Theoretical throughput rates

Processor speed (GHz)	Processor type	Cores	Memory (GBps)	L2 to L3 (GBps)	GX+ (GBps)
1.9	POWER5+	1	21.1	n/a	5.1
1.9	POWER5+	2	21.1	30.4	5.1
2.1	POWER5+	2	21.1	33.6	5.6
1.65	POWER5+	4	21.1	52.8	4.4

2.4 I/O buses

The SCM, DCM, or QCM provides a GX+ bus. In the past, the 6XX bus was the front end from the processor to memory, PCI Host bridge, cache, and other devices. The follow-on of the 6XX bus is the GX bus, which connects the processor to the I/O subsystems. Compared with the 6XX bus, the GX+ bus is both wider and faster and connects to the Enhanced I/O Controller.

The Enhanced I/O Controller is a GX+ to PCI and PCI-X 2.0 Host bridge chip. It contains a GX+ pass-through port and four PCI-X 2.0 buses. The GX+ pass-through port allows other GX+ bus hubs to be connected into the system. Each Enhanced I/O Controller can provide four separate PCI-X 2.0 buses. Each PCI-X 2.0 bus is 64 bits in width and individually capable of running either PCI, PCI-X, or PCI-X 2.0 (DDR only).

Note: The p5-505 has no external RIO-2 ports; and therefore, additional external storage must be attached using other connections, such as a SAN network or SCSI.

2.5 Internal I/O subsystem

PCI-X, where the X stands for extended, is an enhanced PCI bus that delivers a bandwidth of up to 1 GBps, when running a 64-bit bus at 133 MHz or 266 MHz. PCI-X is compatible with earlier systems and can support existing 3.3 volt PCI adapters.

The system provides two full-length PCI-X slots and several integrated I/O devices. Both of these slots are PCI-X DDR and 64-bit capable. The slots run at 266 MHz and are directly connected to the Enhanced I/O Controller. The internal PCI-X slots support a wide range of PCI-X I/O adapters to handle your I/O requirements. The dual 10/100/1000 Mbps Ethernet adapter (two external ports) and the Dual Channel SCSI Ultra320 adapter (a single external port) are examples of integrated devices on the system planar.

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

2.6 64-bit and 32-bit adapters

IBM offers 64-bit adapter options for the p5-505 and p5-505Q, as well as 32-bit adapters. Higher speed adapters use 64-bit slots, because they can transfer 64 bits of data for each data transfer phase. Generally, 32-bit adapters can function in 64-bit PCI-X slots. For a full list of the adapters that are supported in the systems and for important information about adapter placement, see the IBM Systems Hardware Information Center. You can find it at:

http://publib.boulder.ibm.com/infocenter/eserver/v1r3s/index.jsp

2.6.1 LAN adapters

To connect a server to a local area network (LAN), the dual port internal 10/100/1000 Mbps RJ-45 Ethernet controller that is integrated on the system planar can be used.

See Table 2-4 for the list of additional LAN adapters available for an initial system order at the time of writing. IBM supports an installation with NIM using Ethernet and token-ring adapters. The Common Hardware Reference Platform (CHRP), a specification for PowerPC-based systems that can run multiple operating systems, is the platform type.

Table 2-4 Available LAN adapters

Feature code	Adapter description	Type	Slot	Size	Max
1954	4-port 10/100/1000 Ethernet	Copper	32 or 64	short	1
1978	Gigabit Ethernet	Fibre	32 or 64	short	1
1979	Gigabit Ethernet	Copper	32 or 64	short	1
5721	10 Gigabit Ethernet - short reach	Fibre	32 or 64	short	1
5722	10 Gigabit Ethernet - long reach	Fibre	32 or 64	short	1
1983	2-port Gigabit Ethernet	Copper	32 or 64	short	1
1984	2-port Gigabit Ethernet	Fibre	32 or 64	short	1

2.6.2 SCSI adapters

To connect to external SCSI devices, the adapters that are listed in Table 2-5 are available, at the time of writing, for initial order configuration.

Table 2-5 Available SCSI adapters

Feature code	Adapter description	Slot	Size	Max
1912	Dual Channel Ultra320 SCSI	64	short	1
1913	Dual Channel Ultra320 SCSI RAID	64	long	1

Note: Previous SCSI adapters can also be used in the p5-505 or p5-505Q but cannot be part of an initial order configuration. Clients that would like to connect existing external SCSI devices can contact an IBM service representative.

For more information about the internal SCSI system, see 2.7, “Internal storage” on page 32.

2.6.3 Internal RAID option

An option is available to configure internal RAID on a p5-505 or p5-505Q server. The optional SCSI RAID daughter card (FC 1908) plugs directly into the system board to enable this function. FC 1908 is a bootable high performance SCSI RAID feature with RAID 0, 5, or 10 capability. A RAID implementation requires a minimum of three disk drives to form a RAID set.

RAID Capacity limitation: There are limits to the amount of disk drive capacity allowed in a single RAID array. Using the 32-bit AIX 5L kernel, there is a capacity limitation of 1 TB per RAID array. Using the 64-bit kernel, there is a capacity limitation of 2 TB per RAID array. For the RAID adapter and RAID enablement cards, this limitation is enforced by AIX 5L when RAID arrays are created using the PCI-X SCSI Disk Array Manager.

2.6.4 iSCSI

iSCSI is an open, standards-based approach by which SCSI information is encapsulated using the TCP/IP protocol to allow its transport over IP networks. It allows transfer of data between storage and servers in block I/O formats (defined by iSCSI protocol) and thus enables the creation of IP SANs. With iSCSI, an existing network can transfer SCSI commands and data with full location independence and define the rules and processes to accomplish the communication. The iSCSI protocol is defined in iSCSI IETF draft-20.

For more information about this standard, see:

http://tools.ietf.org/html/rfc3720

Although iSCSI can be, by design, supported over any physical media that supports TCP/IP as a transport, today's implementations are only on Gigabit Ethernet. At the physical and link level layers, systems that support iSCSI can be directly connected to standard Gigabit Ethernet switches and IP routers. iSCSI also enables the access to block-level storage that resides on Fibre Channel SANs over an IP network using iSCSI-to-Fibre Channel gateways such as storage routers and switches.

The iSCSI protocol is implemented on top of the physical and data-link layers and presents the operating system with the standard SCSI Access Method command set. It supports SCSI-3 commands and reliable delivery over IP networks. The iSCSI protocol runs on the host initiator and the receiving target device. It can either be optimized in hardware for better performance on an iSCSI host bus adapter (such as FC 1986 and FC 1987 supported in IBM System p5 servers) or run in software over a standard Gigabit Ethernet network interface card. IBM System p5 systems support iSCSI in the following two modes:

Hardware Using iSCSI adapters (see “IBM iSCSI adapters” on page 29).

Software Supported on standard Gigabit adapters, additional software (see “IBM iSCSI software Host Support Kit” on page 29) must be installed. The main processor is utilized for processing related to iSCSI protocol.

Initial iSCSI implementations are targeted for small to medium-sized businesses and departments or branch offices of larger enterprises that have not deployed Fibre Channel SANs. iSCSI is an affordable way to create IP SANs from a number of local or remote storage devices. If there is Fibre Channel present, which is typically present in a data center, it can be accessed by the iSCSI SANs (and vice versa) using iSCSI-to-Fibre Channel storage routers and switches.

iSCSI solutions always involve the following software and hardware components:

Initiators These are the device drivers and adapters that are located on the client. They encapsulate SCSI commands and route them over the IP network to the target device.

Targets The target software receives the encapsulated SCSI commands over the IP network. The software can also provide configuration support and storage-management support. The underlying target hardware can be a storage appliance that contains embedded storage; it can also be a gateway or bridge product that contains no internal storage of its own.

IBM iSCSI adapters

New iSCSI adapters in IBM System p5 systems offer the advantage of increased bandwidth through the hardware support of the iSCSI protocol. The 1 Gigabit iSCSI TOE PCI-X adapters support hardware encapsulation of SCSI commands and data into TCP and transport it over the Ethernet using IP packets. The adapter operates as an iSCSI TCP/IP Offload Engine. This offload function eliminates host protocol processing and reduces CPU interrupts. The adapter uses Small form factor LC type fiber optic connector or copper RJ45 connector.

Table 2-6 lists the iSCSI adapters that can be ordered.

Table 2-6 Available iSCSI adapters

Feature code	Description	Slot	Size	Max
1986	Gigabit iSCSI TOE PCI-X on copper media adapter	64	short	1
1987	Gigabit iSCSI TOE PCI-X on optical media adapter	64	short	1

IBM iSCSI software Host Support Kit

The iSCSI protocol can also be used over standard Gigabit Ethernet adapters. To utilize this approach, download the appropriate iSCSI Host Support Kit for your operating system from the IBM NAS support Web site at:

http://www.ibm.com/storage/support/nas/

The iSCSI Host Support Kit on AIX 5L and Linux acts as a software iSCSI initiator and allows access to iSCSI target storage devices using standard Gigabit Ethernet network adapters. To ensure the best performance, enable TCP Large Send, TCP send and receive flow control, and Jumbo Frame for the Gigabit Ethernet Adapter and the iSCSI target. Tune network options and interface parameters for maximum iSCSI I/O throughput in the operating system.

IBM System Storage N series

The combination of System p5 and IBM System Storage™ N series as the first of a new generation of iSCSI-enabled storage products provides an end-to-end set of solutions. Currently, the System Storage N series features three models: N3700, N5200, and N5500 with:

- Support for entry-level and midrange clients that require Network Attached Storage (NAS) or Internet Small Computer System Interface (iSCSI) functionality
- Support for Network File System (NFS), Common Internet File System (CIFS), and iSCSI protocols
- Data ONTAP software (at no charge), with plenty of additional functions such as data movement, consistent snapshots, and NDMP server protocol, some available through optional licensed functions
- Enhanced reliability with optional clustered (2-node) failover support.

2.6.5 Fibre Channel adapter

The p5-505 and p5-505Q servers support direct or SAN connection to devices using Fibre Channel adapters. Single-port Fibre Channel adapters are available in 2 Gbps and 4 Gbps speeds. A dual-port 4 Gbps Fibre Channel adapter is also available. Table 2-7 provides a summary of the available Fibre Channel adapters.

All of these adapters have LC connectors. If you are attaching a device or switch with an SC type fibre connector an LC-SC 50 Micron Fiber Converter Cable (FC 2456) or an LC-SC 62.5 Micron Fiber Converter Cable (FC 2459) is required.

Supported data rates between the server and the attached device or switch are as follows: Distances of up to 500 meters running at 1 Gbps, distances up to 300 meters running at 2 Gbps data rate, and distances up to 150 meters running at 4 Gbps. When these adapters are used with IBM supported Fibre Channel storage switches supporting long-wave optics, distances of up to 10 kilometers are capable running at 1 Gbps, 2 Gbps, and 4 Gbps data rates.

Table 2-7 Available Fibre Channel adapters

Feature code	Description	Slot	Size	Max
1905	4 Gigabit single-port Fibre Channel PCI-X 2.0 Adapter (LC)	64	short	1
1910	4 Gigabit dual-port Fibre Channel PCI-X 2.0 Adapter (LC)	64	short	1
1977	2 Gigabit Fibre Channel PCI-X Adapter (LC)	64	short	1

2.6.6 Graphic accelerator

The p5-505 and p5-505Q support up to three enhanced POWER GXT135P (FC 1980) 2D graphic accelerators. The POWER GXT135P is a low-priced 2D graphics accelerator for IBM System p5 servers. This adapter supports both analog and digital monitors and is supported for System Management Services (SMS), firmware, and other functions, and when AIX 5L or Linux starts an X11-based GUI.

2.6.7 Asynchronous PCI-X adapters

Asynchronous PCI-X adapters provide a connection for asynchronous EIA-232 or RS-422 devices. In the case of a cluster configuration or high-availability configuration, if the plan is to connect the IBM System p5 servers using a serial connection, the use of the two default system ports is not supported, and you should use one of the features in Table 2-8.

Table 2-8 Asynchronous PCI-X adapters

Feature code	Description
2943	8-Port Asynchronous Adapter EIA-232/RS-422
5723[a]	2-Port Asynchronous IEA-232 PCI Adapter (9-pin)

a. In many cases, the FC 5723 async adapter is configured to supply a backup HACMP heartbeat. In these cases, a serial cable (FC 3927 or FC 3928) must be also configured. Both of these serial cables and the FC 5723 adapter have 9-pin connectors.

2.6.8 PCI-X Cryptographic Coprocessor

The PCI-X Cryptographic Coprocessor (FIPS 4) (FC 4764) for selected System p servers provides both cryptographic coprocessor and secure-key cryptographic accelerator functions in a single PCI-X card. The coprocessor functions are targeted to banking and finance applications. Financial PIN processing and credit card functions are provided. EMV is a standard for integrated chip-based credit cards. The secure-key accelerator functions are targeted to improving the performance of Secure Sockets Layer (SSL) transactions. The FC 4764 provides the security and performance required to support to support On Demand Business and the emerging digital signature application.

The PCI-X Cryptographic Coprocessor (FIPS 4) (FC 4764) provides both cryptographic coprocessor and secure-key cryptographic accelerator functions in a single PCI-X card. The FC 4764 provides secure storage of cryptographic keys in a tamper resistant hardware security module (HSM), which is designed to meet FIPS 140 security requirements. FIPS 140 is a U.S. Government National Institute of Standards & Technology (NIST) administered standard and certification program for cryptographic modules. The firmware for the FC 4764 is available on a separately ordered/distributed CD. This firmware is an LPO product: 5733-CY1 Cryptographic Device Manager. The FC 4764 also requires LPP 5722-AC3 Cryptographic Access Provider to enable data encryption.

Note: This feature has country-specific usage. Refer to the IBM representatives in your country for availability or restrictions.

2.6.9 Additional support for owned PCI-X adapters

The lists of the major PCI-X adapters that can be configured in a system when an initial configuration order is going to be built are described in 2.6.1, “LAN adapters” on page 27 to 2.6.7, “Asynchronous PCI-X adapters” on page 30. The list of all the supported PCI-X adapters, with the related support for additional external devices, is more extended.

Clients that would like to use their own PCI-X adapters can contact the IBM service representative to verify if they are supported.

2.6.10 System ports

The system ports S1 and S2, at the rear of the system, are only available if the system is not managed with an HMC. In this case, the S1 and S2 ports support the attachment of a serial console and modem.

If an HMC is connected, a *virtual serial console* is provided by the HMC (logical device vsa0 for AIX 5L) and a modem can be connected to the HMC. The S1 and S2 ports are not usable in this case.

If serial port function is needed, optional PCI adapters are available, see 2.6.7, “Asynchronous PCI-X adapters” on page 30.

2.6.11 Ethernet ports

The two built-in Ethernet ports provide 10/100/1000 Mbps connectivity over a CAT-5 cable for up to 100 meters. Table 2-9 lists the attributes of the LEDs that are visible on the side of the jack.

Table 2-9 Ethernet LED descriptions

LED	Light	Description
Link	Off Green	No link; could indicate a bad cable, not selected, or configuration error. Connection established.
Activity	On Off	Data activity. Idle.

2.7 Internal storage

One integrated dual-channel Ultra320 SCSI controller, managed by an EADS-X chip, is used to drive the internal disk drives and the external SCSI port. The p5-505 and p5-505Q servers provide two bays that are designed for hot-swappable disk drives. The disk drive backplane docks directly to the system planar. The virtual SCSI Enclosure Services (VSESs) hot-swappable control functions are provided by the integrated Ultra320 SCSI controller. The two internal drives are on SCSI bus 0, which is connected to the internal port on the integrated Ultra320 SCSI controller.

2.7.1 Internal media devices

The p5-505 and p5-505Q servers provide one slim-line media bay for optional DVD drives. Table 2-10 lists available optical media devices. Alternate methods of maintaining and servicing the system must be available if the DVD-ROM or DVD-RAM is not ordered; an external Internet connection must be available to maintain or update system microcode to the latest required level. This control panel/media bay is controlled by the integrated IDE controller.

Table 2-10 Available internal media devices

Feature code	Description
1900	4.7 GB IDE Slimline DVD-RAM Drive
1903	IDE Slimline DVD-ROM Drive

Note: If SUSE Linux Enterprise Server 9 for POWER (or later) or Red Hat Enterprise Linux AS for POWER Version 3 (or later) is being installed in the system. FC 1900, FC 1903, or follow-on is required.

2.7.2 Internal hot-swappable SCSI disks

The p5-505 can have up to two hot-swappable disk drives. The hot-swap process is controlled by the Virtual SCSI enclosure service (SES), which is provided by the integrated SCSI Ultra320 controller. Table 2-11 on page 33 lists available hot-swappable disk drives.

Table 2-11 Hot-swappable disk drive options

Feature code	Description
1970	36.4 GB 15,000 rpm Ultra320 SCSI hot-swappable disk drive
1968	73.4 GB 10,000 rpm Ultra320 SCSI hot-swappable disk drive
1971	73.4 GB 15,000 rpm Ultra320 SCSI hot-swappable disk drive
1969	146.8 GB 10,000 rpm Ultra320 SCSI hot-swappable disk drive
1972	146.8 GB 15,000 rpm Ultra320 SCSI hot-swappable disk drive
1973	300 GB 10,000 rpm Ultra320 SCSI hot-swappable disk drive

The system configuration shipped will have the two SCSI disks installed in DASD slot 1 with SCSI ID 8 and DASD slot 2 with SCSI ID 5. The drive at ID 8 is hardwired to spin up immediately during the startup sequencing. The remaining drive spins up under software control (typically at five second intervals). The disk drive placement priority is SCSI ID 8 and then 5. See Figure 1-2 on page 3 for the SCSI ID location.

Hot-swappable disks and Linux

Hot-swappable disk drives on POWER5 systems are supported with SUSE LINUX Enterprise Server 9 for POWER, or later, and Red Hat Enterprise Linux AS 4 for POWER, or later.

2.8 External disk subsystem

The p5-505 and p5-505Q have internal hot-swappable drives. When the AIX 5L operating system is installed in an IBM System p5 server, the internal disks are normally used for the AIX 5L *rootvg* volume group and paging space. Specific client requirements can be satisfied by the several external disk possibilities that the server supports.

2.8.1 IBM TotalStorage EXP24 Expandable Storage

The IBM TotalStorage EXP24 Expandable Storage disk enclosure, Model D24 or T24, can be purchased together with the p5-505 or p5-505Q and provides low-cost Ultra320 (LVD) SCSI disk storage. This disk storage enclosure device provides more than 7 TB of disk storage in a 4U rack-mount (Model D24) or compact desk side (Model T24) unit. Whether you require high availability storage solutions or simply high capacity storage for a single server installation, the unit offers a cost-effective solution. It provides 24 hot-swappable disk bays with 12 disk bays accessible from the front and 12 disk bays accessible from the rear. Disk options that can be accommodated in any of the four six-pack disk drive enclosures are 73.4 GB, 146.8 GB, or 300 GB 10 K rpm or 36.4 GB, 73.4 GB, or 146.8 GB 15 K rpm drives. Each of the four six-pack disk drive enclosures can be attached independently to an Ultra320 SCSI or Ultra320 SCSI RAID adapter. For high available configurations, a dual bus repeater card (FC 5742) allows each six-pack to be attached to two SCSI adapters that are installed in one or multiple servers or logical partitions. Optionally, the two front or two rear six-packs can be connected together to form a single Ultra320 SCSI bus of 12 drives.

2.8.2 IBM System Storage N3000 and N5000

The IBM System Storage N3000 and N5000 line of iSCSI-enabled storage offerings provide the flexibility for implementing a Storage Area Network over an Ethernet network. The N3000 supports up to 16.8 TB of physical storage and the N5000 supports up to 84 TB of physical disk. Additional information about IBM iSCSI-based storage systems is available at:

http://www.ibm.com/servers/storage/nas/index.html

2.8.3 IBM TotalStorage Storage DS4000 Series

The IBM System Storage DS4000™ line of Fibre Channel-enabled Storage offerings provides a wide range of storage solutions for your SAN. The IBM TotalStorage DS4000 Storage server family consists of the following models: DS4100, DS4300, DS4500, and DS4800. The Model DS4100 Express Product Offering Model is the smallest model and scales up to 44.8 TB; the Model DS4800 is the largest and scales up to 89.6 TB of disk storage at the time of this writing. Model DS4300 provides up to 16 bootable partitions, or 64 bootable partitions if the turbo option is selected, that are attached with the Gigabit Fibre Channel Adapter (FC 1977). Model DS4500 provides up to 64 bootable partitions. Model DS4800 provides 4 GB switched interfaces. In most cases, both the IBM TotalStorage DS4000 family and the IBM System p5 servers are connected to a SAN. If only space for the rootvg is needed, the Model DS4100 is a good solution.

To learn more about the support of additional features and for further information about the IBM TotalStorage DS4000 Storage Server family, refer to the following Web site:

http://www.ibm.com/servers/storage/disk/ds4000/index.html

2.8.4 IBM TotalStorage DS6000 and DS8000 Series

The IBM TotalStorage Enterprise Storage Server® (ESS) DS6000™ and DS8000™ models are the high-end premier storage solution for SANs. They use POWER technology-based design so that they can serve data fast and efficiently.

The IBM TotalStorage DS6000 provides enterprise class capabilities in a space-efficient modular package. It scales to 67.2 TB of physical storage capacity by adding storage expansion enclosures.

The Model DS8000 series is the flagship of the IBM TotalStorage DS family. The DS8000 scales to 192 TB. The DS8000 system architecture is designed to scale to over one petabyte. The Model DS6000 and DS8000 systems can also be used to provide disk space for booting LPARs or partitions using Micro-Partitioning technology. ESS and the IBM System p5 servers are usually connected together to a storage area network.

For further information about ESS, refer to the following Web site:

http://www.ibm.com/servers/storage/disk/enterprise/ds_family.html

2.9 Logical partitioning and virtualization

Dynamic LPARs and virtualization increase the utilization of system resources and add a new level of configuration possibilities. This section provides details and configuration specifications about this topic. The virtualization discussion includes virtualization enabling technologies that are standard in the system, such as the POWER Hypervisor, and optional ones, such as the Advanced POWER Virtualization feature.

2.9.1 Dynamic logical partitioning

Logical partitioning was introduced with the POWER4 processor-based product line and the AIX 5L Version 5.1 operating system. This technology offered the capability to divide a pSeries system into separate logical systems so that each LPAR could run an operating environment on dedicated attached devices, such as processors, memory, and I/O components.

Later, dynamic LPAR increased the flexibility, allowing selected system resources, such as processors, memory, and I/O components, to be added and deleted from dedicated partitions while they are running. AIX 5L Version 5.2, with all the necessary enhancements to enable dynamic LPAR, was introduced in 2002. The ability to reconfigure dynamic LPARs encourages system administrators to redefine all available system resources dynamically to reach the optimum capacity for each defined dynamic LPAR.

Operating system support for dynamic LPAR

Table 2-12 lists AIX 5L and Linux support for dynamic LPAR capabilities.

Table 2-12 Operating system supported function

Function	AIX 5L Version 5.2	AIX 5L Version 5.3	Linux SLES 9	Linux RHEL AS 3	Linux RHEL AS 4
Dynamic LPAR capabilities (add, remove, and move operations)					
Processor	Y	Y	Y	N	Y
Memory	Y	Y	N	N	N
I/O slot	Y	Y	Y	N	Y

2.10 Virtualization

With the introduction of the POWER5 processor, partitioning technology moved from a dedicated resource allocation model to a virtualized shared resource model. This section briefly discusses the key components of virtualization in System p5 servers.

For more information about virtualization, see the following Web site:

http://www.ibm.com/servers/eserver/about/virtualization/systems/pseries.html

See also the following IBM Redbooks:

- *Advanced POWER Virtualization on IBM System p5*, SG24-7940, available at:
  http://www.redbooks.ibm.com/abstracts/sg247940.html?Open
- *Advanced POWER Virtualization on IBM @server p5 Servers: Architecture and Performance Considerations*, SG24-5768, available at:
  http://www.redbooks.ibm.com/abstracts/sg245768.html?Open

2.10.1 POWER Hypervisor

Combined with features that are designed into the POWER5 and POWER5+ processors, the POWER Hypervisor delivers functions that enable other system technologies, including Micro-Partitioning technology, virtualized processors, IEEE virtual local area network (VLAN), compatible virtual switch, virtual SCSI adapters, and virtual consoles. The POWER

Hypervisor is a basic component of system firmware that is always active, regardless of the system configuration.

The POWER Hypervisor also:

- Provides an abstraction between the physical hardware resources and the logical partitions using them
- Enforces partition integrity by providing a security layer between logical partitions
- Controls the dispatch of virtual processors to physical processors (see 2.11.2, “Logical, virtual, and physical processor mapping” on page 39)
- Saves and restores all processor state information during logical processor context switch
- Controls hardware I/O interruption management facilities for LPARs
- Provides virtual LAN channels between physical partitions that help reduce the need for physical Ethernet adapters for communication between partitions

The POWER Hypervisor is always active when the server is running, is partitioned (or is not partitioned), and is not connected to the HMC. It requires memory to support the LPARs on the server. The amount of memory required by the POWER Hypervisor firmware varies according to several factors:

- 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

Note: Use the System Planning Tool for estimate the memory requirements of the POWER Hypervisor.

In AIX 5L V5.3, the `lparstat` command with the `-h` and `-H` flags displays the POWER Hypervisor statistical data. Using the `-h` flag adds summary POWER Hypervisor statistics to the default output of the `lparstat` command.

The minimum amount of physical memory for each partition is 128 MB, but in most cases the actual requirements and recommendations are between 256 MB and 512 MB for AIX 5L, Red Hat, and Novell SUSE Linux. Physical memory is assigned to partitions in increments of Logical Memory Block (LMB). For POWER5+ processor-based systems, LMB can be adjusted from 16 MB to 256 MB.

The following three types of virtual I/O adapters are provided by the POWER Hypervisor.

Virtual SCSI

The POWER Hypervisor provides a virtual SCSI mechanism for virtualization of storage devices (a special LPAR to install the Virtual I/O Server is required to utilize this feature, see 2.11.3, “Virtual I/O Server” on page 41). The storage virtualization is accomplished with two paired adapters: a virtual SCSI server adapter and a virtual SCSI client adapter. Only the Virtual I/O Server partition can define virtual SCSI server adapters; other partitions are *client* partitions. The Virtual I/O Server is available with the optional Advanced POWER Virtualization feature (FC 7432).

Virtual Ethernet

The POWER Hypervisor provides a virtual Ethernet switch function that allows partitions on the same server to communicate quickly and securely without a physical connection. The

virtual Ethernet allows a transmission speed in the range of 1 to 3 GBps depending on the maximum transmission unit (MTU) size and CPU entitlement. Virtual Ethernet requires a system with either AIX 5L Version 5.3 or an appropriate level of Linux supporting virtual Ethernet devices (see 2.13, “Operating system support” on page 50). The virtual Ethernet is part of the base system configuration.

Virtual Ethernet has the following major features:

- The virtual Ethernet adapters can be used for both IPv4 and IPv6 communication and can transmit packets with a size up to 65408 bytes. Therefore, the maximum MTU for the corresponding interface can be up to 65394 (65390 if virtual local area network (VLAN) tagging is used).
- The POWER Hypervisor presents itself to partitions as a virtual 802.1Q-compliant switch. Maximum number of VLANs is 4096. Virtual Ethernet adapters can be configured as either untagged or tagged (following the IEEE 802.1Q VLAN standard).
- A partition supports 256 virtual Ethernet adapters. Besides a default port VLAN ID, the number of additional VLAN ID values that can be assigned per virtual Ethernet adapter is 20, which implies that each virtual Ethernet adapter can be used to access 21 virtual networks.
- Each partition operating system detects the VLAN switch as an Ethernet adapter without the physical link properties and asynchronous data transmit operations.

Any virtual Ethernet can also have connection outside the server if a layer-2 bridging to a physical Ethernet adapter is set in one Virtual I/O server partition (see 2.11.3, “Virtual I/O Server” on page 41 for more details about shared Ethernet).

Note: Virtual Ethernet is based on the IEEE 802.1Q VLAN standard. No physical I/O adapter is required when creating a VLAN connection between partitions and no access to an outside network is required.

Virtual TTY console

Each partition needs to have access to a system console. Tasks such as operating system installation, network setup, and some problem analysis activities require a dedicated system console. The POWER Hypervisor provides the virtual console using a virtual TTY or serial adapter and a set of Hypervisor calls to operate on them. Virtual TTY does not require the purchase of any additional features or software such as the Advanced POWER Virtualization feature.

Depending on the system configuration, the operating system console can be provided by the HMC virtual TTY, Integrated Virtualization Manager virtual TTY, or from a terminal emulator connected to a system port.

2.11 Advanced POWER Virtualization feature

The Advanced POWER Virtualization feature (FC 7432) is an optional, additional cost feature. This feature enables the implementation of more fine-grained virtual partitions on IBM System p5 servers.

The Advanced POWER Virtualization feature includes:

- Firmware enablement for Micro-Partitioning technology

  Support for up to 10 partitions per processor using 1/100 of the processor granularity. Minimum CPU requirement per partition is 1/10. All processors are enabled for

micro-partitions (the number of processors on the system equals the number of Advanced POWER Virtualization features ordered).

- Installation image for the Virtual I/O Server software that is shipped as a system image on DVD. Client partitions can be either AIX 5L V5.3 or Linux. It supports:
  - Ethernet adapter sharing (Ethernet bridge from virtual Ethernet to external network)
  - Virtual SCSI Server
  - Partition management by Integrated Virtualization Manager (Virtual I/O Server V1.2 or later)
- Partition Load Manager (AIX 5L Version 5.3 only) with:
  - Automated CPU and memory reconfiguration
  - Real-time partition configuration and load statistics
  - A GUI

For more details about Advanced POWER Virtualization and virtualization in general, see the following Web site:

http://www.ibm.com/servers/eserver/pseries/ondemand/ve/resources.html

2.11.1 Micro-Partitioning technology

The concept of Micro-Partitioning technology allows you to allocate fractions of processors to the partition. The Micro-Partitioning technology is only available with POWER5 and POWER5+ processor-based systems. From an operating system perspective, a virtual processor cannot be distinguished from a physical processor, unless the operating system has been enhanced to be made aware of the difference. Physical processors are abstracted into virtual processors that are available to partitions. See 2.11.2, “Logical, virtual, and physical processor mapping” on page 39 for more details.

For a shared partition, several options have to be defined:

- Minimum, desired, and maximum processing units. Processing units are defined as processing power, or fraction of time, that the partition is dispatched on physical processors.
- The processing sharing mode, either capped or uncapped.
- Weight (preference) in the case of an uncapped partition.
- Minimum, desired, and maximum number of virtual processors.

POWER Hypervisor calculates a partition’s processing *entitlement* based on its desired processing units and logical processor settings, sharing mode and also based on other active partitions’ requirements. The actual entitlement is never smaller than the desired processing unit’s value and can exceed the desired processing unit’s value if the LPAR is an uncapped partition.

A *partition* can be defined with a processor capacity as small as 0.10 processing units. This represents one-tenth of a physical processor. Each physical processor can be shared by up to 10 processor partitions and the entitlement of a partition can be incremented fractionally by as little as one-hundredth of the processor. The shared processor partitions are dispatched and time-sliced on the physical processors that are under the control of the POWER Hypervisor. The shared processor partitions are created and managed by the HMC or Integrated Virtualization Management (included with Virtual I/O Server software version 1.2 or later). There is only one pool of shared processors at the time of this writing and all shared partitions are dispatched by the Hypervisor in this pool. Dedicated partitions and

micro-partitions can coexist on the same POWER5+ processor-based server as long as enough processors are available.

The server supports up to a 4-core processor configuration; therefore, up to four dedicated partitions, or up to 40 micro-partitions, can be created. It is important to note that the maximums that are stated here are those supported by the hardware, but the practical limits depend on the demands of the application workload.

2.11.2 Logical, virtual, and physical processor mapping

The meaning of the term *physical processor* in this section is a *processor core*. For example, in a 2-core server with a DCM (dual-core module) there are two physical processors.

In dedicated mode, physical processors are assigned as a whole to partitions. The simultaneous multithreading feature in the POWER5+ processor core allows the core to execute instructions from two independent software threads simultaneously. To support this feature, the concept of *logical processors* was introduced. The operating system (AIX 5L or Linux) sees one physical processor as two logical processors if the simultaneous multithreading feature is turned on. It can be turned off while the operating system is executing (for AIX 5L, use the **smtctl** command). If simultaneous multithreading is off, then each physical processor is presented as one logical processor, and thus only one thread is executed on the physical processor at a time.

In a micro-partitioned environment with shared mode partitions, an additional concept, *virtual processors,* was introduced. Shared partitions can define any number of virtual processors (maximum number is 10 times the number of processing units assigned to the partition). From the POWER Hypervisor point of view, the virtual processors represent dispatching objects (for example, the POWER Hypervisor dispatches virtual processors to physical processors according to the partition's processing units entitlement). At the end of the POWER Hypervisor's dispatch cycle (10 ms), all partitions should receive total CPU time equal to their processing units entitlement. Virtual processors are either running (dispatched) on a physical processor or standby (waiting). An operating system is able to dispatch its software threads to these virtual processors and is completely screened from the actual number of physical processors. The logical processors are defined on top of virtual processors in the same way as though they are physical processors. So, even with a virtual processor, the concept of logical processor exists and the number of logical processors depends on whether the simultaneous multithreading is turned on or off.

The following additional information is related to virtual processors:

- There is one-to-one mapping of running virtual processors to physical processors at any given time. The number of virtual processors that can be active at any given time cannot exceed the total number of physical processors in the shared processor pool.
- A virtual processor can be either running (dispatched) on a physical processor or standby and waiting for a physical processor to become available.
- Virtual processors do not introduce any additional abstraction level, they are really only a dispatch entity. When running on a physical processor, virtual processors run at the same speed as the physical processor.
- Each partition's profile defines the CPU entitlement that determines how much processing power any given partition should receive. The total sum of CPU entitlement of all partitions cannot exceed the number of available physical processors in the shared processor pool.
- A partition has the same amount of processing power regardless of the number of virtual processors that it defines.

- A partition can use more processing power, regardless of its entitlement, if it is defined as an *uncapped* partition in the partition profile. If there is spare processing power available in the shared processor pool or other partitions are not using their entitlement, an uncapped partition can use additional processing units if its entitlement is not enough to satisfy its application processing demand in the given processing entitlement.
- When the partition is uncapped, the number of defined virtual processors determines the limitation of the maximum processing power it can receive. For example, if the number of virtual processors is two, then the maximum usable processor units are two.
- You are allowed to define more virtual processors than physical processors. In that case, the virtual processor will be waiting for dispatch more often and some performance impact caused by redispatching virtual processors on physical processors should be considered. It is also true that some applications can benefit from using more virtual processors than the physical processors.
- The number of virtual processors can be changed dynamically through a dynamic LPAR operation.

Virtual processor recommendations

For each partition, you can define a number of virtual processors set to the maximum processing power that the partition can request. If there are, for example, four physical processors installed in the system, one production partition and three test partitions, then:

- Define the production LPAR with four virtual processors so that it can receive full processing power from all four physical processors during the time that the other partitions are idle.
- If you know that the test system is never going to consume more than one processor computing unit, then the test system should be defined with one virtual processor. Some test systems might require additional virtual processors, such as four, so that they can use idle processing power left over by a production system during off-business hours.

Figure 2-7 shows logical, virtual, and physical processor mapping, and an example of how the virtual processor and logical processor can be dispatched to the physical processor.

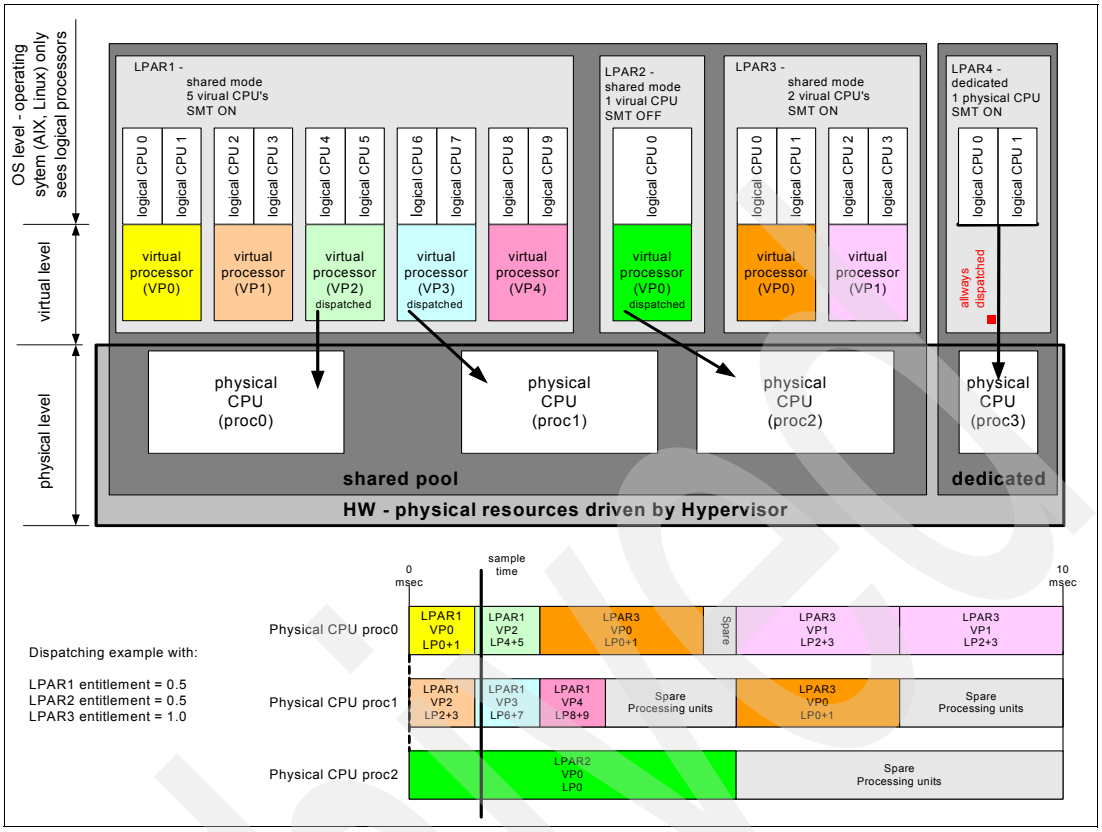


Figure 2-7 Logical, virtual, and physical processor mapping

In Figure 2-7, a system with four physical processors and four partitions is presented; one partition (LPAR4) is in dedicated mode and three partitions (LPAR1, LPAR2, and LPAR3) are running in shared mode. Dedicated mode LPAR4 is using one physical processor and therefore three processors are available for that shared processor pool. LPAR1 defines five virtual processors and the simultaneous multithreading feature is on (so that LPAR1 sees 10 logical processors), LPAR2 defines one virtual processor and simultaneous multithreading is off (one logical processor). LPAR3 defines two virtual processors and simultaneous multithreading is on. Currently (sample time), virtual processors 2 and 3 of LPAR1 and virtual processor 0 of LPAR2 are dispatched on physical processors in the shared pool. Other virtual processors are idle, waiting for dispatch by Hypervisor. When more virtual processors are defined in a partition, any virtual processor shares equal parts of partition processing entitlement.

2.11.3 Virtual I/O Server

The Virtual I/O Server (VIOS) is a special purpose partition that provides virtual I/O resources to other partitions. The Virtual I/O Server owns the physical resources (SCSI, Fibre Channel, network adapters, and optical devices) and allows client partitions to share access to them, which minimizes the number of physical adapters in the system. The Virtual I/O Server eliminates the requirement for every partition to own a dedicated network adapter, disk adapter, and disk drive.

Figure 2-8 shows the organization of a micro-partitioned system, including the Virtual I/O Server. The system also has virtual SCSI and Ethernet connections and mixed operating system partitions.

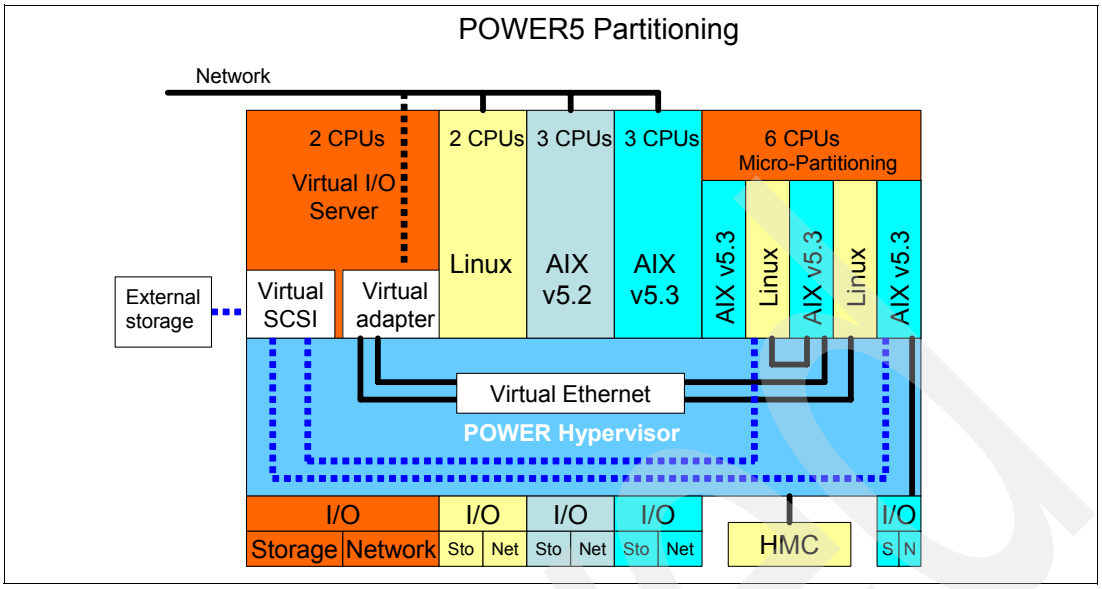


Figure 2-8 Micro-Partitioning technology and VIOS

Because the Virtual I/O Server is an operating system-based appliance server, redundancy for physical devices attached to the Virtual I/O Server can be provided with capabilities such as Multipath I/O and IEEE 802.3ad Link Aggregation.

Installation of the Virtual I/O Server partition is performed from a special system backup DVD that is provided to clients who order the Advanced POWER Virtualization feature. This dedicated software is only for the Virtual I/O Server (and Integrated Virtualization Manager if it is used) and is only supported in special Virtual I/O Server partitions.

The Virtual I/O Server can be installed with:

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

Note: To increase the performance of I/O-intensive applications, use dedicated physical adapters using dedicated partitions.

We recommend that you install the Virtual I/O Server in a partition with dedicated resources or at least 0.5 processor entitlement to help ensure consistent performance.

The Virtual I/O Server supports RAID configurations and SAN-attached devices (possibly with a multipath driver). Logical volumes that are created on RAID or JBOD configurations are bootable, and the number of logical volumes is limited to the amount of storage available and the architectural limits of the Logical Volume Manager.

Two major functions are provided with the Virtual I/O Server: a shared Ethernet adapter and Virtual SCSI.

Shared Ethernet adapter

A shared Ethernet adapter is a Virtual I/O Server service that acts as a layer 2 network bridge between a physical Ethernet adapter or aggregation of physical adapters (EtherChannel) and

one or more virtual Ethernet adapters that are defined by the Hypervisor on the Virtual I/O Server. With a shared Ethernet adapter, LPARs on the virtual Ethernet can share access to the physical Ethernet and communicate with stand-alone servers and LPARs on other systems. The shared Ethernet network provides this access by connecting the internal Hypervisor VLANs with the VLANs on the external switches. Because the shared Ethernet network processes packets at layer 2, the original MAC address and VLAN tags of the packet are visible to other systems on the physical network. IEEE 802.1 VLAN tagging is supported.

The virtual Ethernet adapters that are used to configure a shared Ethernet adapter must have the trunk setting enabled. The trunk setting causes these virtual Ethernet adapters to operate in a special mode so that they can deliver and accept external packets from the POWER5 internal switch to the external physical switches. The trunk setting should only be used for the virtual Ethernet adapters that are part of a shared Ethernet network setup in the Virtual I/O Server.

A single shared Ethernet adapter setup can have up to 16 virtual Ethernet trunk adapters and each virtual Ethernet trunk adapter can support up to 20 VLAN networks. Therefore, it is possible for a single physical Ethernet to be shared among 320 internal VLANs. The number of shared Ethernet adapters that can be set up in a Virtual I/O Server partition is limited only by the resource availability, because there are no configuration limits.

For a more detailed discussion about virtual networking, see:

`http://www.ibm.com/servers/aix/whitepapers/aix_vn.pdf`

Virtual SCSI

Access to real storage devices is implemented through the virtual SCSI services, a part of the Virtual I/O Server partition. This is accomplished with a pair of virtual adapters: a virtual SCSI server adapter and a virtual SCSI client adapter. The virtual SCSI server and client adapters are configured with an HMC or through Integrated Virtualization Manager on smaller systems. The virtual SCSI server (target) adapter is responsible for executing any SCSI commands that it receives. It is owned by the Virtual I/O Server partition. The virtual SCSI client adapter allows a client partition to access physical SCSI-attached and SAN-attached devices and LUNs that are assigned to the client partition.

Physical disks that are owned by the Virtual I/O Server partition can either be exported and assigned to a client partition as a whole device, or can be configured into a volume group and partitioned into several logical volumes. These logical volumes can then be assigned to individual partitions. From the perspective of client partitions, these two options are equivalent.

The Virtual I/O Server provides mapping between *backing devices* (physical devices or logical volumes that are assigned to client partitions in VIOS nomenclature) and client partitions by a command line interface. The appropriate command is the **`mkvdev`** command. For syntax and semantics, see Virtual I/O Server documentation.

All current storage device types, such as SAN, SCSI, and RAID are supported. SSA and iSCSI are not supported at the time of writing.

For more information about the specific storage devices supported, see:

`http://techsupport.services.ibm.com/server/vios/home.html`

Important: We do not recommend using Mirrored Logical Volumes (LVs) on the Virtual I/O Server level as backing devices. If mirroring is required, two independent devices (possibly from two separate Virtual I/O (VIO) servers) should be assigned to the client partition and the client partition should define mirroring on top of them.

Virtual I/O Server version 1.3

Virtual I/O Server version 1.3 brings a host of new enhancements including improved monitoring such as additional **topas** and **viostat** performance metrics and the bundling of the Performance ToolKit (PTX®) agent. Virtual SCSI and virtual Ethernet performance increases, command line enhancements, and enablement of additional storage solutions are also included.

Virtual I/O Server 1.3 introduced several enhancements for Virtual SCSI and shared Fiber Channel adapter support:

- Independent Software Vendor/Independent Hardware Vendor Virtual I/O enablement
- iSCSI TOE adapter
- iSCSI direct-attached n3700 storage subsystem
- HP storage
- Virtual SCSI functional enhancements:
  - Support for SCSI Reserve/Release for limited configurations
  - Changeable queue depth
  - Updating virtual device capacity non-disruptively so that the virtual disk can "grow" without requiring a reconfiguration
  - Configurable fast fail time (number of retries on failure)
  - Error log enhancements

Virtual I/O Server 1.3 also introduced several enhancements for virtual Ethernet and shared Ethernet adapter support, including TCP/IP Acceleration: Large Block Send.

2.11.4 Partition Load Manager

Partition Load Manager provides automated processor and memory distribution between a dynamic LPAR and a Micro-Partitioning technology-capable logical partition that is running AIX 5L. The Partition Load Manager application is based on a client/server model to share system information, such as processor or memory events, throughout the concurrent present logical partitions.

The following events are registered on all managed partition nodes:

- Memory-pages-steal high thresholds and low thresholds
- Memory-usage high thresholds and low thresholds
- Processor-load-average high threshold and low threshold

Note: Partition Load Manager is supported on AIX 5L Version 5.2 and AIX 5L Version 5.3. It is not supported on Linux.

2.11.5 Integrated Virtualization Manager

In order to ease virtualization technology adoption in any IBM System p5 environment, IBM has developed Integrated Virtualization Manager (IVM) — a simplified hardware management solution that inherits some HMC features, thus avoiding the necessity of a dedicated control workstation. This solution enables the administrator to reduce system setup time. IVM is targeted at small and medium systems.

IVM supports up to the maximum 16-core configuration. The IVM provides a management model for a single system. Although it does not provide the full flexibility of an HMC, it enables the exploitation of the IBM Virtualization Engine™ technology. IVM is an enhancement of the Virtual I/O Server offered as part of Virtual I/O Server Version 1.2 and follow-on versions, which is the product that enables I/O virtualization in POWER5 and POWER5+ systems. It provides the same Virtual I/O Server features plus a Web-based graphical interface that enables the administrator to remotely manage the System p5 server with an Internet browser.

Integrated Virtualization Manager can be used to complete the following tasks:

- Create and manage logical partitions.
- Configure the virtual Ethernet networks.
- Manage storage in the Virtual I/O Server.
- Create and manage user accounts.
- Create and manage serviceable events through Service Focal Point.
- Download and install updates to device microcode and to Virtual I/O Server software.
- Back up and restore logical partition configuration information.
- View application logs and the device inventory.

The requirements for an Integrated Virtualization Manager-managed server are as follows:

- A server managed by Integrated Virtualization Manager cannot be simultaneously managed by an HMC.
- Integrated Virtualization Manager (with Virtual I/O Server) must be installed as the first operating system.
- An Integrated Virtualization Manager partition requires a minimum of one virtual processor and 512 MB of RAM.

Virtual I/O Server Version 1.3 introduced enhancements to IVM. The Integrated Virtualization Manager (IVM) adds an industry leading function in this release: support for Dynamic Logical Partitioning (DLPAR) for memory and processors in managed partitions. Additionally, a number of usability enhancements include support through the browser-based interface for IP configuration of the Virtual I/O Server:

- DLPAR support for memory and processors in managed partitions
- GUI support for System Plan management, including the Logical Partition (LPAR) Deployment Wizard
- Web User Interface (UI) support for:
  - IP configuration support
  - Task Manager for long-running tasks
  - Various usability enhancements, including the ability to create a new partition based on an existing one

The major considerations of Integrated Virtualization Manager in comparison to an HMC-managed system are as follows:

- All physical adapters are owned by Integrated Virtualization Manager, and LPARs use virtual devices.
- There is only one profile per partition.
- A maximum of four virtual Ethernet networks are available inside the system.
- Each LPAR can have a maximum of one Virtual SCSI adapter assigned.

- IVM supports a single Virtual I/O Server to support all your mission critical production needs.
- Service Agent (see 3.2.3, “Service Agent” on page 73) for reporting hardware errors to IBM is not available when using the Integrated Virtualization Manager.
- Integrated Virtualization Manager cannot be used by HACMP software to activate Capacity on Demand (CoD) resources on machines that support CoD.

Integrated Virtualization Manager provides advanced virtualization functionality without the need for an extra-cost workstation. For more information about Integrated Virtualization Manager functionality and best practices, see *Virtual I/O Server Integrated Virtualization Manager*, REDP-4061:

http://www.ibm.com/systems/p/hardware/meetp5/ivm.pdf

Figure 2-9 on page 46 shows how a system with Integrated Virtualization Manager is organized. There is a Virtual I/O server and Integrated Virtualization Manager installed in one partition that owns all physical server resources and four client partitions. Integrated Virtualization Manager communicates to the POWER Hypervisor to *create*, *manage,* and *provide virtual I/O* for client partitions. But the dispatch of partitions on physical processors is done by the POWER Hypervisor as in HMC-managed servers. The rules for mapping the physical processors, virtual processors, and logical processors apply as discussed in 2.11.2, “Logical, virtual, and physical processor mapping” on page 39 for shared partitions that are managed by the HMC.

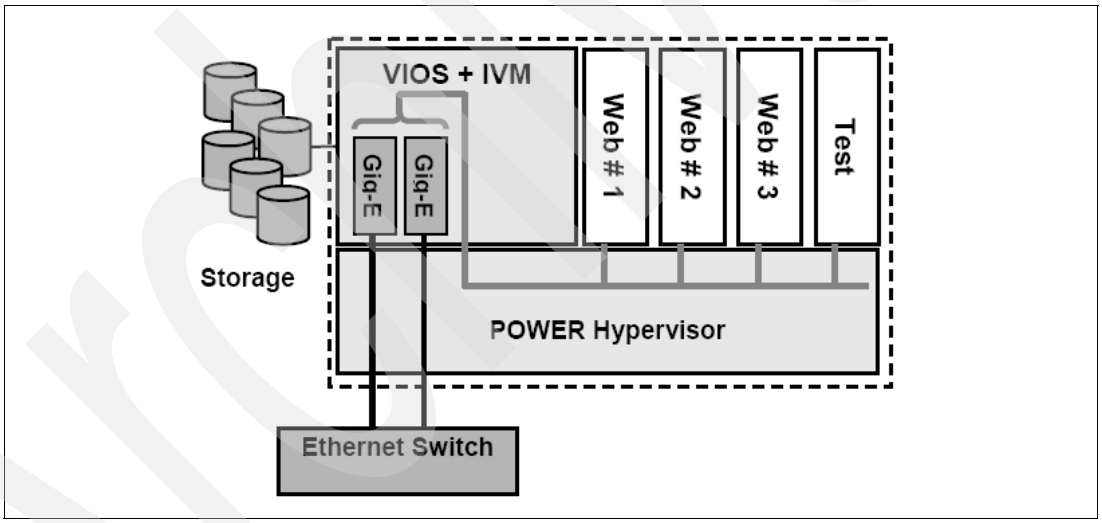


Figure 2-9 Integrated Virtualization Manager principles

Note: Integrated Virtualization Manager and HMC are two separate management systems and cannot be used at the same time; Integrated Virtualization Manager targets cost of ownership, while HMC targets flexibility and scalability. The internal design is so different that no HMC must ever be connected to a working Integrated Virtualization Manager system. If a client wants to migrate an environment from Integrated Virtualization Manager to HMC, the configuration setup has to be manually rebuilt.

Operating system support for advanced virtualization

Table 2-13 lists AIX 5L and Linux support for advanced virtualization.

Table 2-13 Operating system supported functions

Advanced POWER Virtualization feature	AIX 5L Version 5.2	AIX 5L Version 5.3	Linux SLES 9	Linux RHEL AS 3	Linux RHEL AS 4
Micro-partitions (1/10th of processor)	N	Y	Y	Y	Y
Virtual Storage	N	Y	Y	Y	Y
Virtual Ethernet	N	Y	Y	Y	Y
Partition Load Manager	Y	Y	N	N	N

2.12 Hardware Management Console

The HMC is a dedicated workstation that provides a graphical user interface for configuring, operating, and performing basic system tasks for the System p5 servers in either non-partitioned, LPAR, or clustered environments. In addition, the HMC is used to configure and manage partitions. One HMC is capable of controlling multiple POWER5 and POWER5+ processor-based systems.

At the time of writing, one HMC supports up to 48 POWER5 and POWER5+ processor-based systems and up to 254 LPARs using the HMC machine code Version 5.1. For updates of the machine code and HMC functions and hardware prerequisites, refer to the following Web site:

https://www14.software.ibm.com/webapp/set2/sas/f/hmc/home.html

POWER5 and POWER5+ processor-based system HMCs require Ethernet connectivity between HMC and the service processor of the server; moreover, if dynamic LPAR operations are required, all AIX 5L and Linux partitions must be enabled to communicate over the network to HMC. Ensure that sufficient Ethernet adapters are available to enable public and private networks, if you need both:

- The HMC 7310 Model C05 is a desk side model with one integrated 10/100/1000 Mbps Ethernet port and two additional PCI slots.
- The 7310 Model CR3 is a 1U, 19-inch rack-mountable drawer that has two native 10/100/1000 Mbps Ethernet ports and two additional PCI slots.

For any partition in a server, it is possible to use the shared Ethernet adapter in the Virtual I/O Server for a unique connection from the HMC to partitions. Therefore, client partitions do not require their own physical adapters to communicate with HMC.

It is a good practice to connect the HMC to the first HMC port on the system, labeled as HMC Port 1; although other network configurations are possible. A second HMC can be attached to HMC Port 2 of the server for redundancy (or vice versa). Figure 2-10 shows a simple network configuration for the connection from HMC to server and Dynamic LPAR operations. For more details about HMC and the possible network connections, refer to *Hardware Management Console (HMC) Case Configuration Study for LPAR Management*, REDP-3999, which is at the following Web site:

http://www.redbooks.ibm.com/abstracts/redp3999.html

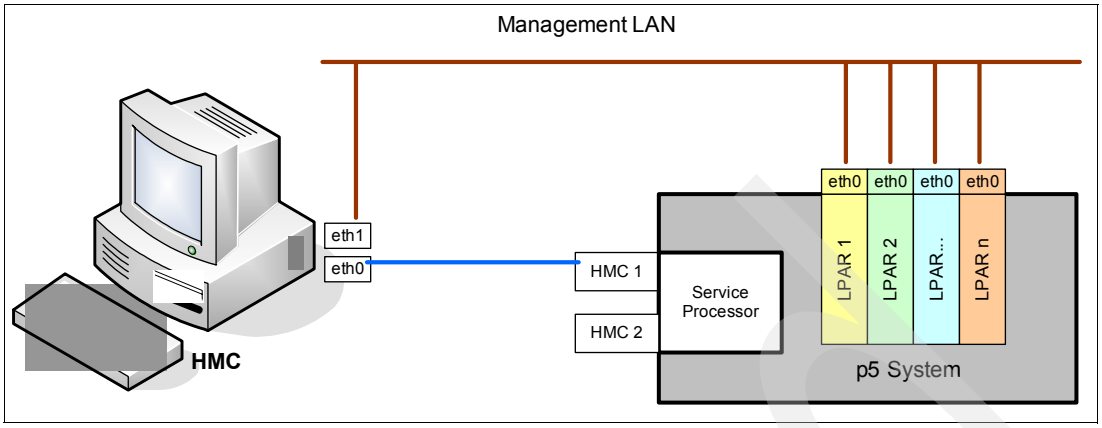


Figure 2-10 HMC to service processor and LPARs network connection

The default mechanism for the allocation of the IP addresses for the service processor HMC ports is dynamic. The HMC can be configured as a DHCP server, providing the IP address at the time that the managed server is powered on. If the service processor of the managed server does not receive the DHCP reply before timeout, predefined IP addresses will set up on both ports. Static IP address allocation is also an option. You can configure the IP address of the service processor ports with a static IP address with the Advanced System Management Interface (ASMI) menus. See 2.14.7, “Service processor” on page 61 for predefined IP addresses and additional information.

Note: If you have to access ASMI (for example, to set up the IP address of a new POWER5+ processor-based server when HMC is not available or not providing DHCP services), you can connect any client to one of the service processor HMC ports with any kind of Ethernet cable, and use a Web browser to access the predefined IP address, such as the following example:

`https://192.168.2.147`

The functions performed by the HMC include:

- Creating and maintaining a multiple partition environment
- Displaying a virtual operating system session terminal for each partition
- Displaying a virtual operator panel of contents for each partition
- Detecting, reporting, and storing changes in hardware conditions
- Powering managed systems on and off
- Acting as a service focal point

The HMC provides a graphical and a command-line interface for all management tasks. Remote connection to the HMC using Web-based System Manager or SSH is possible. For accessing the graphical interface, you can use the Web-based System Manager Remote Client running on AIX 5L, Linux, or Windows®. The Web-based System Manager client installation image can be downloaded from the HMC itself from the following URL:

`http://<hmc_address_or_name>/remote_client.html`

Both unencrypted and encrypted Web-based System Manager connections are supported. The command line interface is also available with the SSH secure shell connection to the HMC. It can be used by an external management system or a partition to perform HMC operations remotely.

2.12.1 High availability using the HMC

The HMC is an important hardware component. HACMP Version 5.3 High Availability cluster software can be used to automatically activate resources (where available), thereby becoming an integral part of the cluster. For some environments, working with redundant HMCs is recommended.

POWER5 and POWER5+ processor-based systems have two service processor interfaces (HMC port 1 and HMC port 2) that available for connections to the HMC. We recommend that you use both of them for redundant network configuration. Depending on your environment, you have multiple options to configure the network. Figure 2-11 shows one possible highly available configuration.

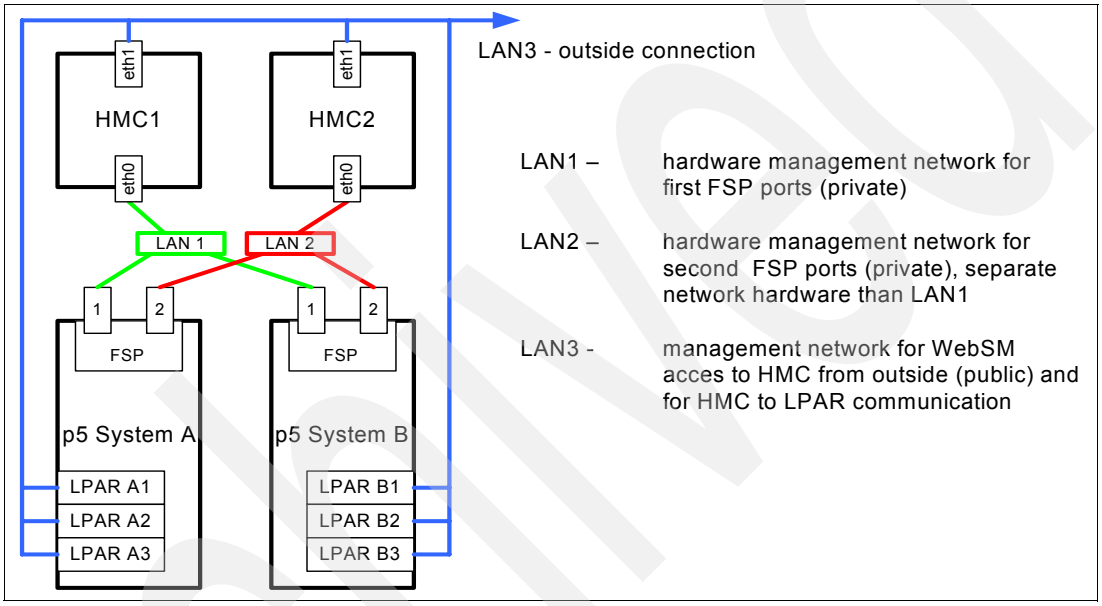


Figure 2-11 Highly available HMC and network architecture

Note that only the hardware management network (LAN1 and LAN2) is highly available for the sake of simplicity. However, the management network (LAN3) can be made highly available using a similar concept and adding more Ethernet adapters to LPARs and HMCs.

2.12.2 IBM System Planning Tool

The IBM System Planning Tool (SPT) is the next generation of the IBM LPAR Validation Tool (LVT). It contains all of the function from the LVT and is integrated with the IBM Systems Workload Estimator (WLE). System plans generated by the SPT can be deployed on the system by the Hardware Management Console (HMC). The SPT is available to assist the user in system planning, design, validation, and to provide a system validation report that reflects the user's system requirements while not exceeding system recommendations. The SPT is a PC-based browser application designed to be run in a stand-alone environment.

You can download the IBM System Planning Tool at no additional charge from:

`http://www.ibm.com/servers/eserver/support/tools/systemplanningtool/`

The System Planning Tool (SPT) helps you design a system to fit your needs. You can use the SPT to design a logically partitioned system or you can use the SPT to design an unpartitioned system. You can create an entirely new system configuration, or you can create a system configuration based upon any of the following:

- Performance data from an existing system that the new system is to replace
- Performance estimates that anticipate future workloads that you must support
- Sample systems that you can customize to fit your needs

Integration between the SPT and both the Workload Estimator (WLE) and IBM Performance Management (PM) allows you to create a system that is based upon performance and capacity data from an existing system or that is based on new workloads that you specify.

You can use the SPT before you order a system to determine what you must order to support your workload. You can also use the SPT to determine how you can partition a system that you already have.

Important: We recommend using the IBM System Planning Tool to estimate Hypervisor requirements and determine the memory resources that are required for all partitioned and non-partitioned servers.

Figure 2-12 on page 50 shows the estimated Hypervisor memory requirements based on sample partition requirements.

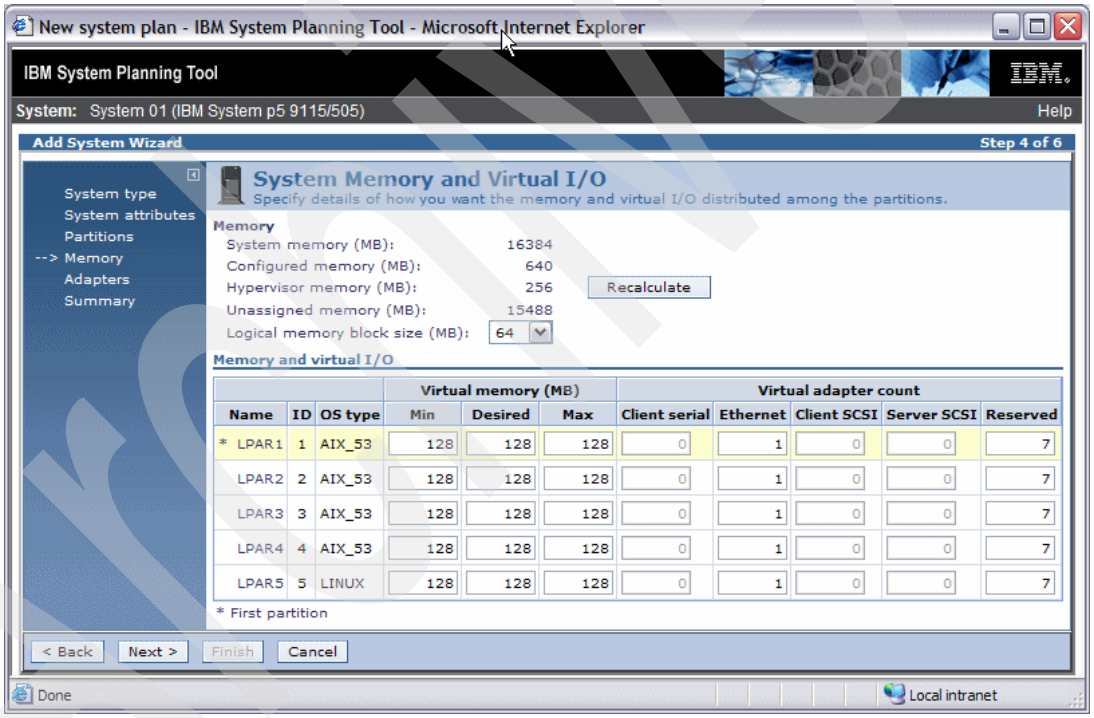


Figure 2-12 IBM System Planning Tool window showing Hypervisor requirements

2.13 Operating system support

The p5-505 and p5-505Q are capable of running AIX 5L and Linux. AIX 5L has been specifically developed and enhanced to exploit and support the extensive RAS features on System p servers.

2.13.1 AIX 5L

If installing AIX 5L on the p5-505 or p5-505Q, you must meet the following minimum requirements:

- POWER5+ Processors:
  - AIX 5L for POWER V5.2 with the 5200-09 Technology Level (APAR IY82425), or later
  - AIX 5L for POWER V5.3 with the 5300-05 Technology Level (APARIY82426), or later
- POWER5 Processors:
  - AIX 5L V5.2 with the 5200-07 Recommended Maintenance Package (APAR IY67914), or later
  - AIX 5L V5.3 with the 5300-03 Recommended Maintenance Package (APAR IY71011), or later

Note: The Advanced POWER Virtualization feature (FC 7432) is not supported on AIX 5L V5.2; it requires AIX 5L V5.3.

IBM periodically releases maintenance packages for the AIX 5L operating system. These packages are available on CD-ROM or you can download them from the Internet at:

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

The Web page provides information about how to obtain the CD-ROM.

You can also get individual operating system fixes and information about obtaining AIX 5L service at this Web site. In AIX 5L V5.3, the **suma** command is also available, which helps the administrator automate the task of checking and downloading operating system downloads. For more information about the **suma** command functionality, refer to:

http://www14.software.ibm.com/webapp/set2/sas/f/suma/home.html

If you have problems downloading the latest maintenance level, ask your IBM Business Partner or IBM representative for assistance.

AIX 5L is also available on DVD. Table 2-14 lists the order numbers.

Table 2-14 Order numbers for AIX 5L media

Order number	Description	Media
LCD4-7544-00	AIX 5L V5.3 Base media Maintenance Level 3	DVD
LCD4-7549-00	AIX 5L V5.2 Base media Maintenance Level 7	DVD

Electronic Software Delivery (ESD) for AIX 5L V5.2 and V5.3 for POWER5 systems was made available. This is a way for clients to receive software and associated publications online versus waiting for a physical shipment to arrive. Clients requesting ESD should order new FC 3450.

ESD has the following requirements:

- POWER5 system
- Internet connectivity from a POWER5 system or PC
- Connectivity speeds greater than 56 Kbps for downloading large products such as AIX 5L
- Registration on the ESD Web site

For additional information, contact your IBM sales representative.

2.13.2 Linux

For the p5-505 and p5-505Q, Linux distributions are available through Novell SUSE and Red Hat at the time of writing this publication. The p5-505 requires the following version of Linux distributions:

- SUSE LINUX Enterprise Server 9 for POWER, or later
- Red Hat Enterprise Linux AS 4 for POWER, or later

Note: Not all p5-505 features available on AIX 5L are available on Linux.

For information about the features and external devices supported by Linux on the p5-505, refer to:

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

For information about SUSE LINUX Enterprise Server 9, refer to:

http://www.novell.com/products/linuxenterpriseserver/

For information about Red Hat Enterprise Linux AS, refer to:

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

Many of the features described in this document are operating system dependent and might not be available on Linux. For more information, see:

http://www-03.ibm.com/systems/p/software/whitepapers/linux_overview.pdf

Note: IBM only supports the Linux systems of clients with a SupportLine contract covering Linux. Otherwise, contact the Linux distributor for support.

New discounted Linux subscriptions

Linux subscriptions are now available when ordered through IBM and combined with an IBM System p5 Express Product Offering configuration. Clients can purchase a one-year discounted subscription or a greater discount for a three-year subscription.

These new Linux options, available on System p5 Express Product Offering servers, bring improved pricing and price performance to our clients interested in Linux as their primary operating system. Clients interested in AIX 5L can also obtain an Express Product Offering that fits their needs.

Clients are still encouraged to purchase support for their Linux subscription either though IBM Global Services or through the distributor to receive updates and technical assistance as needed. Support is not included in the price of the subscription.

The new lower-priced Linux subscriptions, when combined with the lower package prices of the System p5 Express Product Offerings, make these products an exceptional value for our smaller to mid-market clients, as well as larger enterprises.

Refer to the following Web site for Red Hat information:

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

For additional information about Linux running on OpenPower systems, visit:

http://www.ibm.com/servers/eserver/openpower/

For additional information about Linux on POWER, visit:

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

2.14 Service information

The p5-505 is a client setup server and is shipped with materials to assist in the general installation of the server. The server cover has a quick reference service information label that provides graphics that can aid in identifying features and location information. This section provides some additional service-related information.

2.14.1 Touch point colors

Blue (IBM blue) or terra-cotta (orange) on a component indicates a *touch point* (for electronic parts) where you can grip the hardware to remove it from or install it into the system, open or close a latch, and so on. IBM defines the touch point colors as follows:

Blue This requires a shutdown of the system before the task can be performed, for example, removing the PCI riser book to install PCI adapters in the p5-505.

Terra-cotta The system can remain powered on while this task is being performed. Keep in mind that some tasks might require that other steps be performed first. One example is deconfiguring a physical volume in the operating system before removing the disk from the p5-505.

Blue and terra-cotta Terra-cotta takes precedence over this color combination, and the rules for a terra-cotta-only touch point apply.

Important: It is important to adhere to the touch point colors on the system. Not doing so can compromise your safety and damage the system.

2.14.2 Securing a system into a rack

The *optional* rack-mount drawer rail kit is a unique kit designed for use with the p5-505. No tools are required to install the p5-505 or drawer rails into the system rack.

The kit has a modular design that can be adapted to accommodate various rack depth specifications. The drawer rails are equipped with thumb-releases on the sides, toward the front of the server, that allow for easy slide out from its rack position for servicing.

Attention: Always exercise standard safety precautions when installing or removing devices from racks.

To place or slide the p5-505 or p5-505Q in the service position:

1. If necessary, open the front rack door.
2. If they are present, remove the two thumbscrews, *A*, that secure the server unit to the rack, as shown in Figure 2-13 on page 54.
3. Release the rack latches, *B*, on both the left and right sides, as shown in the same figure.
4. Review the following notes, and then slowly pull the server unit out from the rack until the rails are fully extended and locked:

– If the procedure that you are performing requires you to unplug cables from the back of the server, do so before you pull the unit out from the rack.
– Ensure that the cables at the rear of the system unit do not catch or bind as you pull the server out from the system rack.
– When the rails are fully extended, the rail safety latches lock into place. This action prevents the server from being pulled out too far.

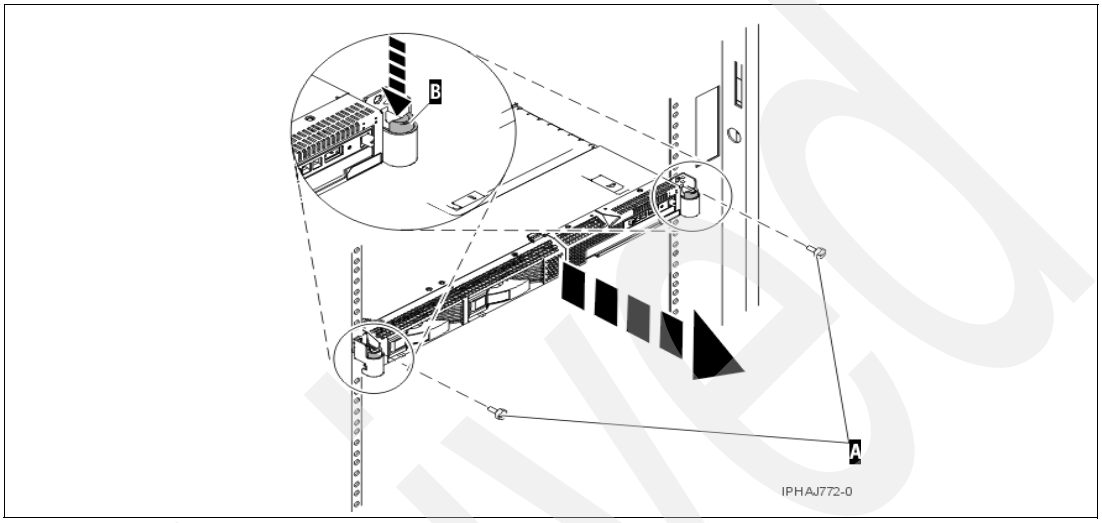


Figure 2-13 Pull the p5-505 to the service position

5. Continue to release the p5-505 or p5-505Q out of the rack. Press the rail safety latches, *A*, to release the server from the system rack, as shown in Figure 2-14.

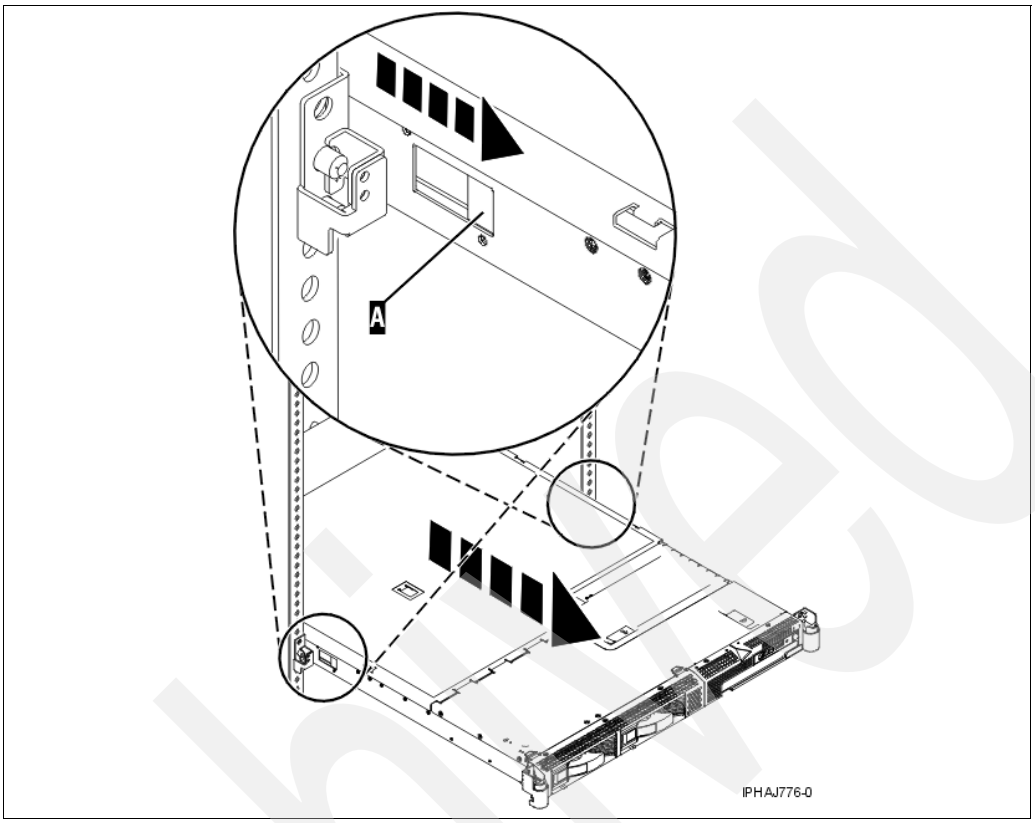


Figure 2-14 Release the p5-505 or p5-505Q out of the rack

6. Grasp each side of the server unit and pull the server out of the system rack.

> **Attention:** This unit weighs approximately 17 kg (37 pounds). Ensure that you can safely support this weight when removing the server unit from the system rack.

7. Place the server on a sturdy flat surface capable of safely supporting the server while you are servicing it.

2.14.3 Fault identification button

After the service person powers off the p5-505 or p5-505Q and removes it from the rack, there is a button in the middle of the server planar that lights the LEDs of any failed component that is on the planar. This process is accomplished by pushing the button once. The power used for these diagnostics is provided by an internal battery pack. This feature is based on client feedback that indicated the need to reaffirm the location of a failed component after the system is powered down and safely disconnected from an external power source.

The arrow in Figure 2-15 points to the fault identification button.

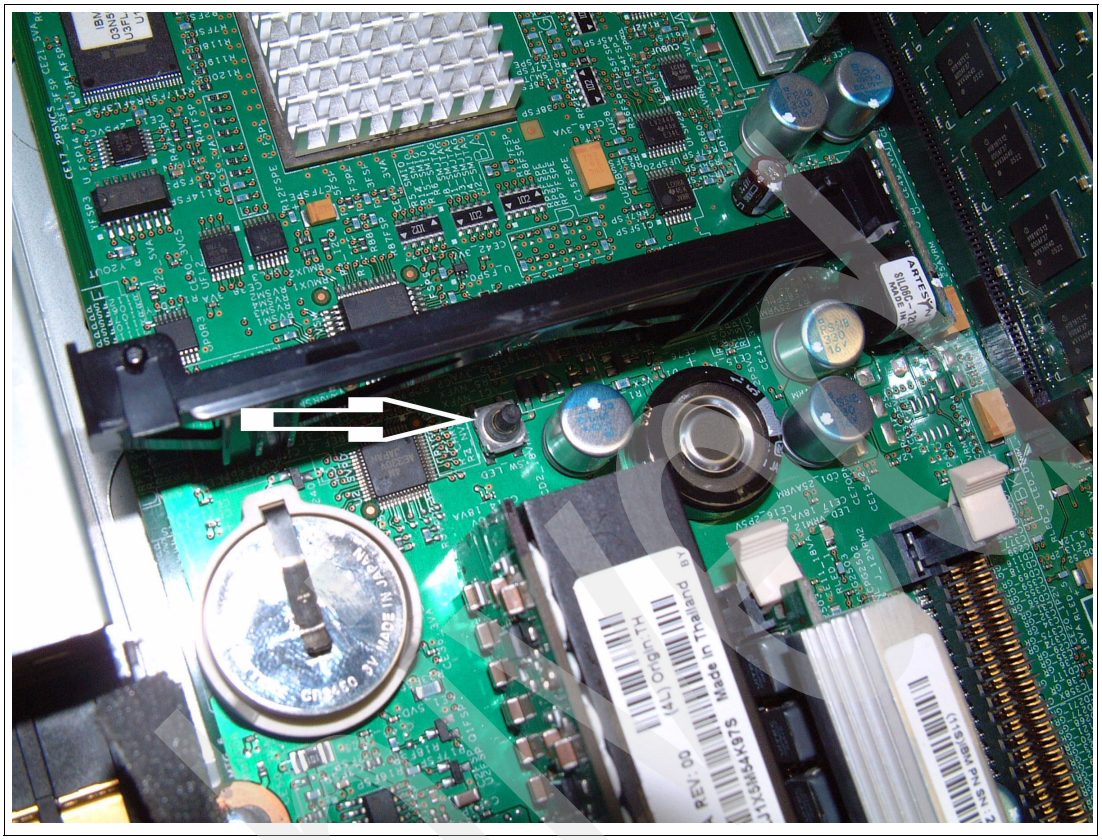


Figure 2-15 Fault identification button

For more information, see the IBM eServer pSeries and AIX 5L Information Center:

http://publib.boulder.ibm.com/infocenter/eserver/v1r3s/index.jsp

2.14.4 Operator control panel

The service processor provides an interface to the control panel that is used to display server status and diagnostic information. The p5-505 and p5-505Q control panels are packaged so that they fit into a smaller space. In the normal position, the control panel is seated inside the chassis on the right top of the DVD optical device (if viewed from the front of the server). The LCD display is invisible from the front. To read the LCD display, the client or engineer needs to pull the operator panel out toward the front.

Note: For servers managed by the HMC, use it to perform control panel functions.

Accessing the p5-505 or p5-505Q control panel

To access all of the control panel's features, perform the following steps (refer to Figure 2-16):

1. Press inward on the spring-loaded tab, *B*, located on the right side of the control panel, *B*, so that it pops out slightly.
2. Pull out the control panel toward the front of the server until it can be pivoted downward on its hinge.

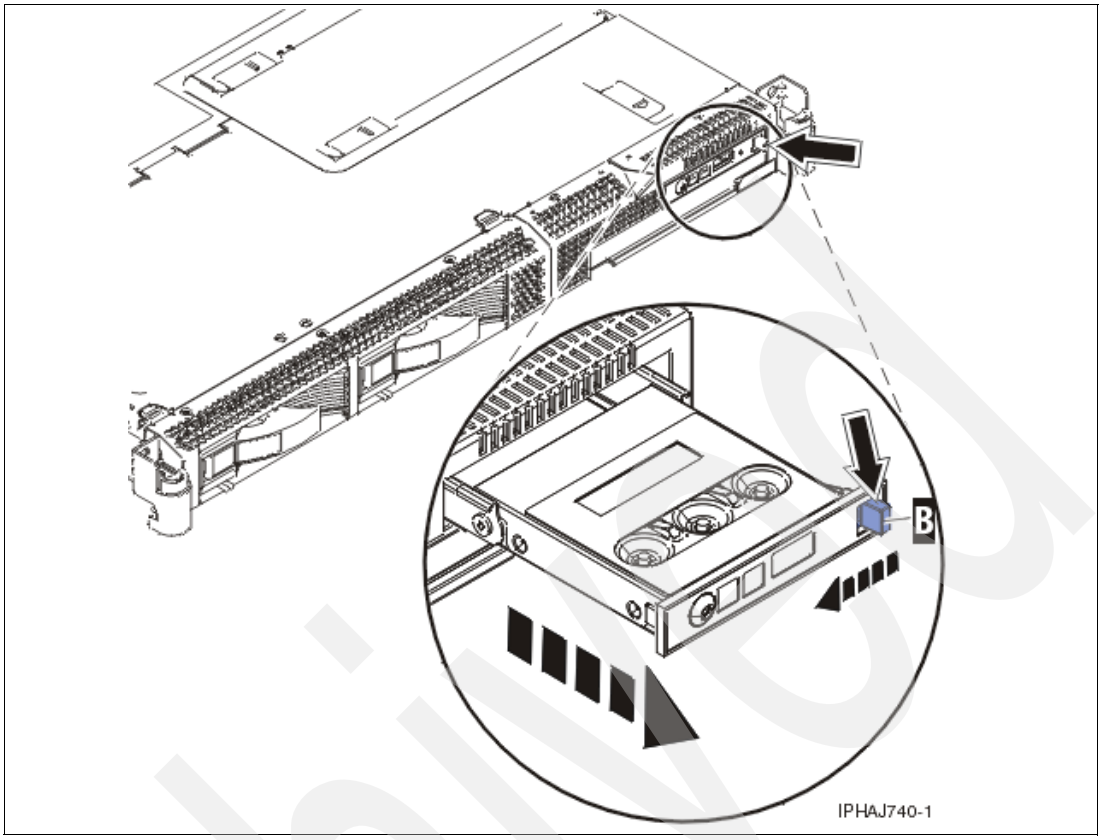


Figure 2-16 Accessing the p5-505 or p5-505Q control panel

3. To move the control panel back into the device enclosure, lift the control panel up to align it with the opening and push it into place until you feel the tab lock.

Primary control panel functions

The primary control panel functions are defined as functions 01 to 20, including options to view and manipulate IPL modes, server operating modes, IPL speed, and IPL type.

The following list describes all of the available primary functions:

- Function 01: Display selected IPL type, system operating mode, and IPL speed
- Function 02: Select IPL type, IPL speed override, and system operating mode
- Function 03: Start IPL
- Function 04: Lamp Test
- Function 05: Reserved
- Function 06: Reserved
- Function 07: SPCN functions
- Function 08: Fast Power Off
- Functions 09 to 10: Reserved
- Functions 11 to 19: System Reference Code
- Function 20: System type, model, feature code, and IPL type

All of the functions mentioned are accessible using the Advanced System Management Interface (ASMI), HMC, or the control panel.

For detailed information about each control panel function and the available values, select **Service provider information** → **Reference information** → **Service functions** → **Control panel functions** from the IBM Systems Hardware Information Center Web site at:

http://publib.boulder.ibm.com/infocenter/eserver/v1r3s/index.jsp?lang=en

2.14.5 Cable-management arm

The p5-505 and p5-505Q are shipped with a cable-management arm. To install the cable-management arm, refer to Figure 2-17 and follow these steps:

1. From the back of the system rack, locate the cable-management arm flange, *A*, located on the fixed back portion of the left server rail assembly.

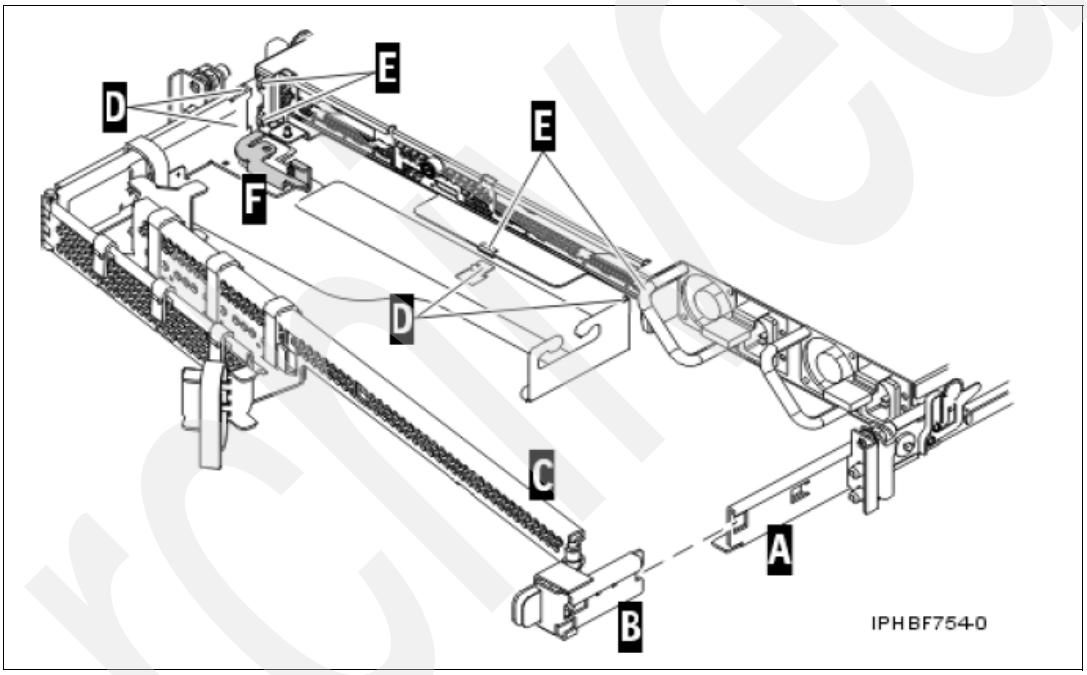


Figure 2-17 Cable-management arm and system unit

2. Attach the cable-management arm clasp, *B*, to the rail by pushing the clasp onto the rail until it locks into place.
3. Attach the other end of the cable-management arm to the back of the server by performing the following steps:
   a. Align the tabs, *D*, on the cable-management arm with the slots, *E*, on the back of the server.
   b. Slide the cable-management arm to the left, securing it in place. Make sure that all the tabs fit into the slots.
   c. Push the locking lever, *F*, into the locked position.

Note: Ensure that the cable-management arm, *C*, is level so that it moves freely.

2.14.6 System firmware

Server firmware is the part of the Licensed Internal Code that enables hardware, such as the service processor. Depending on your service environment, you can download, install, and manage your server firmware fixes using different interfaces and methods, including the HMC, or by using functions specific to your operating system.

Note: Normally, installing the server firmware fixes through the operating system is a nonconcurrent process.

Temporary and permanent firmware sides

The service processor maintains two copies of the server firmware:

- One copy is considered the permanent or backup copy and is stored on the permanent side, sometimes referred to as the *"p" side*.
- The other copy is considered the installed or temporary copy and is stored on the temporary side, sometimes referred to as the *"t" side*. We recommend that you start and run the server from the temporary side.

The copy actually booted from is called the *activated level*, sometimes referred to as *"b"*.

Note: The default value, from which the system boots, is temporary.

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

- AIX 5L:

```
The current permanent system firmware image is SF220_005.
The current temporary system firmware image is SF220_006.
The system is currently booted from the temporary image.
```

- Linux:

```
system:SF220_006 (t) SF220_005 (p) SF220_006 (b)
```

When you install a server firmware fix, it is installed on the temporary side.

Note: The following points are of special interest:

- The server firmware fix is installed on the temporary side only after the existing contents of the temporary side are permanently installed on the permanent side (the service processor performs this process automatically when you install a server firmware fix).
- If you want to preserve the contents of the permanent side, you need to remove the current level of firmware (copy the contents of the permanent side to the temporary side) before you install the fix.
- However, if you get your fixes using the Advanced features on the HMC interface and you indicate that you do not want the service processor to automatically accept the firmware level, the contents of the temporary side are not automatically installed on the permanent side. In this situation, you do not need to remove the current level of firmware to preserve the contents of the permanent side before you install the fix.

You might want to use the new level of firmware for a period of time to verify that it works correctly. When you are sure that the new level of firmware works correctly, you can

permanently install the server firmware fix. When you permanently install a server firmware fix, you copy the temporary firmware level from the temporary side to the permanent side.

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

For a detailed description of firmware levels, select **Customer service, support, and troubleshooting** → **Fixes and upgrades** → **Getting fixes and upgrades** from the IBM Systems Hardware Information Center Web site at:

http://publib.boulder.ibm.com/infocenter/eserver/v1r3s/index.jsp?lang=en

System firmware download Web site

For the system firmware download Web site for the p5-505, go to:

http://www14.software.ibm.com/webapp/set2/firmware/gjsn

Receive server firmware fixes using an HMC

If you use an HMC to manage your server and you periodically configure several partitions on the server, you need to download and install fixes for your server and power subsystem firmware.

How you get the fix depends on whether the HMC or server is connected to the Internet:

- The HMC or server is connected to the Internet.

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

- Neither the HMC nor your server is connected to the Internet (server firmware only).

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

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

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

For a detailed description of each task, select **Customer service, support, and troubleshooting** → **Fixes and upgrades** → **Getting fixes and upgrades** from the IBM Systems Hardware Information Center Web site at:

http://publib.boulder.ibm.com/infocenter/eserver/v1r3s/index.jsp?lang=en

Receive server firmware fixes without an HMC

Periodically, you need to install fixes for your server firmware. If you do not use an HMC to manage your server, you must get your fixes through your operating system. In this situation, you can get server firmware fixes through the operating system regardless of whether your operating system is AIX 5L or Linux.

To do this, complete the following tasks:

1. Determine the existing level of server firmware using the **lsmcode** command.
2. Determine the available levels of server firmware.
3. Get the server firmware.
4. Install the server firmware fix to the temporary side.
5. Verify that the server firmware fix installed successfully.
6. Install the server firmware fix permanently (optional).

Note: To view existing levels of server firmware using the **lsmcode** command, you need to have the following service tools installed on your server:

- AIX 5L

  You must have AIX 5L diagnostics installed on your server to perform this task. AIX 5L diagnostics are installed when you install AIX 5L on your server. However, it is possible to deselect the diagnostics. Therefore, you need to ensure that the online AIX 5L diagnostics are installed before proceeding with this task.

- Linux
  - Platform Enablement Library: librtas-*nnnnn*.rpm
  - Service Aids: ppc64-utils-*nnnnn*.rpm
  - Hardware Inventory: lsvpd-*nnnnn*.rpm

    Where *nnnnn* represents a specific version of the RPM file.

  If you do not have the service tools on your server, you can download them at the following Web site:

  `http://www14.software.ibm.com/webapp/set2/sas/f/lopdiags`

2.14.7 Service processor

The service processor is an embedded controller running the service processor internal operating system. The service processor operating system contains specific programs and device drivers for the service processor hardware. The host interface is a 32-bit PCI-X interface connected to the Enhanced I/O Controller.

The service processor is used to monitor and manage the system hardware resources and devices. The service processor offers connections through two Ethernet 10/100 Mbps ports:

- Both Ethernet ports are only visible to the service processor and can be used to attach the server to an HMC or to access the Advanced System Management Interface (ASMI) options from a client Web browser, using the HTTP-server integrated into the service processor internal operating system.
- Both Ethernet ports have a default IP address:
  - Service processor Eth0 or HMC1 port is configured as 192.168.2.147 with netmask 255.255.255.0.

– Service processor Eth1 or HMC2 port is configured as 192.168.3.147 with netmask 255.255.255.0.

For the major functions of service processor, see 3.2.1, “Service processor” on page 71.

2.14.8 Hardware management user interfaces

In this section, we provide a brief overview of the different p5-505 or p5-505Q hardware management user interfaces available.

Advanced System Management Interface

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

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

You might be able to use the service processor's default settings. In that case, accessing the ASMI is not necessary.

Accessing the ASMI using a Web browser

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

Accessing the ASMI using an ASCII console

The Advanced System Management Interface on an ASCII console supports a subset of the functions provided by the Web interface and is available only when the system is in the platform standby state. The ASMI on an ASCII console is not available during some phases of system operation, such as the initial program load and run time.

Accessing the ASMI using an HMC

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

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

System Management Services

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

To start System Management Services, perform the following steps:

1. For a server that is connected to an HMC, use the HMC to restart the server or partition.

   If the server is not connected to an HMC, stop the system, and then restart the server by pressing the power button on the control panel.

2. For a partitioned server, watch the virtual terminal window on the HMC.

   For a full server partition, watch the firmware console.

3. Look for the Power-on self-test (POST) indicators for memory, keyboard, network, scsi, and speaker that appear across the bottom of the screen. Press the numeric 1 key after the word `keyboard` appears and before the word `speaker` appears.

HMC

The Hardware Management Console is a system that controls managed systems, including IBM System p hardware, logical partitions, and Capacity on Demand. To provide flexibility and availability, there are different ways to implement HMCs, including a local HMC, remote HMC, redundant HMC, and the Web-based System Manager Remote Client.

Local HMC

A local HMC is any physical HMC that is directly connected to the server it manages through a private service network. An HMC in a private service network is a Dynamic Host Control Protocol (DHCP) server from which the managed server obtains the address for its firmware. Additional local HMCs in your private service network are DHCP clients.

Remote HMC

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

Redundant HMC

A redundant HMC manages a server that is already managed by another HMC. When two HMCs manage one server, those HMCs are peers and can be used simultaneously to manage the server. The redundant HMC in your private service network is usually a DHCP client.

Web-based System Manager Remote Client

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

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

For more detailed information about the use of the HMC, refer to the IBM Systems Hardware Information Center.

Open Firmware

A System p5 server has one instance of Open Firmware both when in the partitioned environment and when running as a full system partition. Open Firmware has access to all devices and data in the server. Open Firmware is started when the server goes through a power-on reset. Open Firmware, which runs in addition to the Hypervisor in a partitioned environment, runs in two modes: global and partition. Each mode of Open Firmware shares the same firmware binary that is stored in the flash memory.

In a partitioned environment, Open Firmware runs on top of the global Open Firmware instance. The partition Open Firmware is started when a partition is activated. Each partition has its own instance of Open Firmware and has access to all the devices assigned to that partition. However, each instance of Open Firmware has no access to devices outside of the partition in which it runs. Partition firmware resides within the partition memory and is replaced when AIX 5L or Linux takes control. Partition firmware is needed only for the time that is necessary to load AIX 5L or Linux into the partition server memory.

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

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

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

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

3

Reliability, availability, and serviceability

This chapter provides information about IBM System p5 design features that help lower the total cost of ownership (TCO). IBM reliability, availability, and serviceability (RAS) technology allow you to improve your TCO architecture by reducing unplanned down time. This chapter includes several features based on the benefits that are available when using AIX 5L. Support of these features using Linux can vary.

3.1 Reliability, fault tolerance, and data integrity

Excellent quality and reliability are inherent in all aspects of the IBM System p5 processor design and manufacturing. The fundamental objective of the design approach is to minimize outages. The RAS features help to ensure that the system operates when required, performs reliably, and efficiently handles any failures that might occur. This is achieved using capabilities that are provided by both the hardware and the operating system AIX 5L.

The p5-505 or p5-505Q as a POWER5+ server enhances the RAS capabilities that are implemented in POWER4-based systems. RAS enhancements available on POWER5 and POWER5+ servers are:

- Most firmware updates allow the system to remain operational.
- The ECC has been extended to inter-chip connections for the fabric and processor bus.
- Partial L2 cache deallocation is possible.
- The number of L3 cache line deletes improved from two to ten for better self-healing capability.

The following sections describe the concepts that form the basis of leadership RAS features of IBM System p5 systems in more detail.

3.1.1 Fault avoidance

IBM System p5 servers are built on a quality-based design that is intended to keep errors from happening. This design includes the following features:

- Reduced power consumption and cooler operating temperatures for increased reliability, which is enabled by the use of copper circuitry, silicon-on-insulator, and dynamic clock gating
- Mainframe-inspired components and technologies

3.1.2 First-failure data capture

If a problem should occur, the ability to diagnose that problem correctly is a fundamental requirement upon which improved availability is based. The p5-505 and p5-505Q incorporate advanced capability in start-up diagnostics and in run-time First-failure data capture (FDDC) based on strategic error checkers built into the processors.

Any errors detected by the pervasive error checkers are captured into Fault Isolation Registers (FIRs), which can be interrogated by the service processor. The service processor has the capability to access system components using special purpose ports or by access to the error registers. Figure 3-1 on page 67 shows a schematic of a Fault Register Implementation.

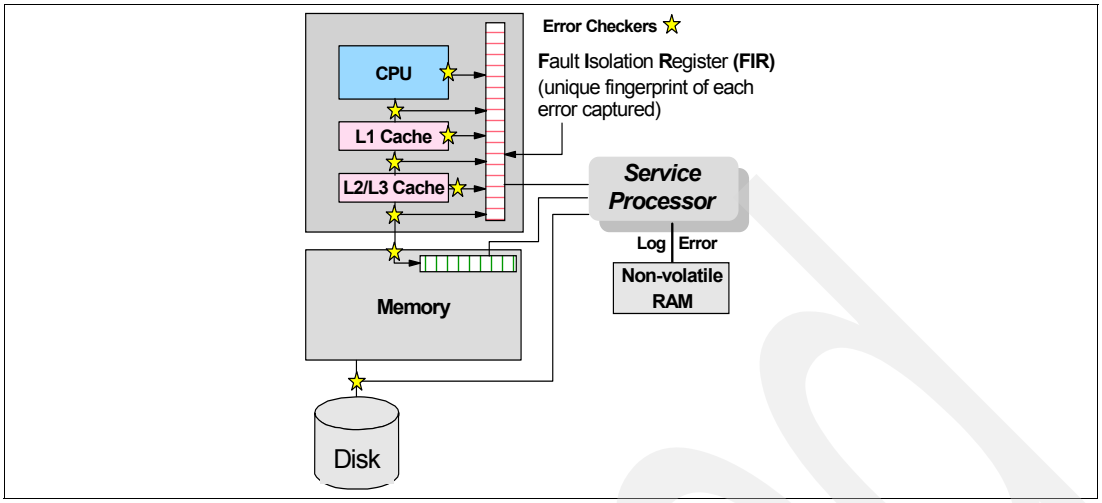


Figure 3-1 Schematic of Fault Isolation Register implementation

The FIRs are important because they enable an error to be uniquely identified, thus enabling the appropriate action to be taken. Appropriate actions might include such things as a bus retry, ECC correction, or system firmware recovery routines. Recovery routines can include dynamic deallocation of potentially failing components.

Errors are logged into the system non-volatile random access memory (NVRAM) and the service processor event history log, along with a notification of the event to AIX 5L for capture in the operating system error log. Diagnostic Error Log Analysis (*diagela*) routines analyze the error log entries and invoke a suitable action such as issuing a warning message. If the error can be recovered, or after suitable maintenance, the service processor resets the FIRs so that they can record any future errors accurately.

The ability to correctly diagnose any pending or firm errors is a key requirement before any dynamic or persistent component deallocation or any other reconfiguration can take place.

For further details, see 3.1.7, “Resource deallocation” on page 69.

3.1.3 Permanent monitoring

The service processor (SP) included in the p5-505 or p5-505Q provides a way to monitor the system even when the main processor is inoperable.

Mutual surveillance

The SP can monitor the operation of the firmware during the boot process, and it can monitor the operating system for loss of control. This allows the service processor to take appropriate action, including calling for service, when it detects that the firmware or the operating system has lost control. Mutual surveillance also allows the operating system to monitor for service processor activity and can request a service processor repair action if necessary.

Environmental monitoring

Environmental monitoring related to power, fans, and temperature is done by the System Power Control Network (SPCN). Environmental critical and non-critical conditions generate Early Power-Off Warning (EPOW) events. Critical events (for example, Class 5 ac power loss) trigger appropriate signals from hardware to impacted components in order to prevent any data loss without operating system or firmware involvement. Non-critical environmental events are logged and reported using Event Scan.

The operating system cannot program or access the temperature threshold using the SP.

EPOW events can, for example, trigger the following actions:

- Temperature monitoring, which increases the fan's speed rotation when ambient temperature is above a preset operating range.
- Temperature monitoring warns the system administrator of potential environmental-related problems. It also performs an orderly system shutdown when the operating temperature exceeds a critical level.
- Voltage monitoring provides warning and an orderly system shutdown when the voltage is out of the operational specification.

3.1.4 Self-healing

For a system to be self-healing, it must be able to recover from a failing component by first detecting and isolating the failed component, taking it offline, fixing or isolating it, and reintroducing the fixed or replacement component into service without any application disruption. Examples include:

- *Bit steering* to redundant memory in the event of a failed memory chip 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 using Error Checking and Correcting (ECC) without reaching error thresholds for main, L2, and L3 cache memory
- L3 cache line deletes extended from 2 to 10 for additional self-healing
- ECC extended to inter-chip connections on fabric and processor bus
- *Memory scrubbing* to help prevent soft-error memory faults

Memory reliability, fault tolerance, and integrity

The p5-505 and p5-505Q use Error Checking and Correcting (ECC) circuitry for system memory to correct single-bit and to detect double-bit memory failures. Detection of double-bit memory failures helps maintain data integrity. Furthermore, the memory chips are organized such that the failure of any specific memory chip only affects a single bit within a four-bit ECC word (*bit-scattering*), thus allowing for error correction and continued operation in the presence of a complete chip failure (*Chipkill recovery*). The memory DIMMs also use *memory scrubbing* and thresholding to determine when spare memory chips within each bank of memory should be used to replace ones that have exceeded their threshold of error count (*dynamic bit-steering*). Memory scrubbing is the process of reading the contents of the memory during idle time and checking and correcting any single-bit errors that have accumulated by passing the data through the ECC logic. This function is a hardware function on the memory controller and does not influence normal system memory performance.

3.1.5 N+1 redundancy

The use of redundant parts allows the p5-505 and p5-505Q to remain operational with full resources:

- Redundant spare memory bits in L1, L2, L3, and main memory
- Redundant fans
- Redundant power supplies (optional)

Important: With this optional feature, every rack-mount p5-505 or p5-505Q requires two power cords, which are not included in the base order. For maximum availability, we highly recommend that you connect power cords from the same p5-505 or p5-505Q to two separate PDUs in the rack. These PDUs should be connected to two independent client power sources.

3.1.6 Fault masking

If corrections and retries succeed and do not exceed threshold limits, the system remains operational with full resources, and no intervention is required:

- CEC bus retry and recovery
- PCI-X bus recovery
- ECC Chipkill soft error

3.1.7 Resource deallocation

If recoverable errors exceed threshold limits, resources can be deallocated with the system remaining operational, allowing deferred maintenance at a convenient time.

Dynamic or persistent deallocation

Dynamic deallocation of potentially failing components is nondisruptive, allowing the system to continue to run. Persistent deallocation occurs when a failed component is detected, which is then deactivated at a subsequent reboot.

Dynamic deallocation functions include:

- Processor
- L3 cache line delete
- Partial L2 cache deallocation
- PCI-X bus and slots

For dynamic processor deallocation, the service processor performs a predictive failure analysis based on any recoverable processor errors that have been recorded. If these transient errors exceed a defined threshold, the event is logged and the processor is deallocated from the system while the operating system continues to run. This feature (named *CPU Guard*) enables maintenance to be deferred until a suitable time. Processor deallocation can only occur if there are sufficient functional processors (at least two).

To verify whether CPU Guard has been enabled, run the following command:

```
lsattr -El sys0 | grep cpuguard
```

If enabled, the output will be similar to the following:

```
cpuguard    enable      CPU Guard     True
```

If the output shows CPU Guard as disabled, enter the following command to enable it:

```
chdev -l sys0 -a cpuguard='enable'
```

Cache or cache-line deallocation is aimed at performing dynamic reconfiguration to bypass potentially failing components. This capability is provided for both L2 and L3 caches. Dynamic run-time deconfiguration is provided if a threshold of L1 or L2 recovered errors is exceeded.

In the case of an L3 cache run-time array single-bit solid error, the spare resources are used to perform a line delete on the failing line.

PCI hot-plug slot fault tracking helps prevent slot errors from causing a system machine check interrupt and subsequent reboot. This provides superior fault isolation, and the error affects only the single adapter. Run-time errors on the PCI bus caused by failing adapters result in recovery action. If this is unsuccessful, the PCI device is shut down gracefully. Parity errors on the PCI bus itself result in bus retry, and if uncorrected, the bus and any I/O adapters or devices on that bus are deconfigured.

The p5-505 or p5-505Q supports PCI Extended Error Handling (EEH) if it is supported by the PCI-X adapter. In the past, PCI bus parity errors caused a global machine check interrupt, which eventually required a system reboot in order to continue. In the p5-505 or p5-505Q system, hardware, system firmware, and AIX 5L interaction have been designed to allow transparent recovery of intermittent PCI bus parity errors and graceful transition to the I/O device available state in the case of a permanent parity error in the PCI bus.

EEH-enabled adapters respond to a special data packet generated from the affected PCI slot hardware by calling system firmware, which examines the affected bus, allows the device driver to reset it, and continues without a system reboot.

Persistent deallocation functions include:

- Processor
- Memory
- Deconfigure or bypass failing I/O adapters
- L3 cache

Following a hardware error that has been flagged by the service processor, the subsequent reboot of the system invokes extended diagnostics. If a processor or L3 cache is marked for deconfiguration by persistent processor deallocation, the boot process attempts to proceed to completion with the faulty device deconfigured automatically. Failing I/O adapters are deconfigured or bypassed during the boot process.

Note: The auto-restart (reboot) option, when enabled, can reboot the system automatically following an unrecoverable software error, software hang, hardware failure, or environmentally induced failure (such as loss of power supply).

3.1.8 Serviceability

Increasing service productivity means the system is up and running for a longer time. The p5-505 and p5-505Q improve service productivity by providing the functions described in the following sections.

Error indication and LED indicators

Thep5-505 and p5-505Q are designed for client setup of the machine and for the subsequent addition of most hardware features. The p5-505 and p5-505Q also allow clients to replace service parts (Client Replaceable Unit). To accomplish this, the p5-505 and p5-505Q provide internal LED diagnostics that identify parts that require service. Attenuation of the error is provided through a series of light attention signals, starting on the exterior of the system (System attention LED) located on the front of the system, and ending with an LED near the failing Field Replaceable Unit.

For more information about Client Replaceable Units, including videos, see:

http://publib.boulder.ibm.com/infocenter/eserver/v1r3s/index.jsp

System attention LED

The attention indicator is represented externally by an amber LED on the operator panel and the back of the system unit. It is used to indicate that the system is in one of the following states:

- Normal state, LED is off.
- Fault state, LED is on solid.
- Identify state, LED is blinking.

Additional LEDs on I/O components such as PCI-X slots and disk drives provide status information such as power, hot-swap, and the need for service.

Concurrent maintenance

Concurrent maintenance provides replacement of the following parts while the system remains running:

- Disk drives
- Cooling fans
- Power subsystems
- PCI-X adapter cards

3.2 Manageability

The functions and tools provided for IBM System p5 servers to ease management are described in the next sections.

3.2.1 Service processor

The service processor (SP) is always working. The CEC can be in the following states:

- Power standby mode (power off)
- Operating, ready to start partitions
- Operating with some partitions running and an AIX 5L or Linux system in control of the machine

The SP is still working and checking the system for errors, ensuring the connection to the HMC (if present) for manageability purposes and accepting Advanced System Management Interface (ASMI) SSL network connections. The SP provides the possibility to view and manage the machine-wide settings using the ASMI and allows complete system and partition management from the HMC. Also, the surveillance function of the SP is monitoring the operating system to check that it is still running and has not stalled.

Note: The IBM System p5 service processor enables the analysis of a system that does not boot. It can be performed either from the ASMI, the HMC, or the ASCI console (depending on the presence of the HMC). ASMI is provided in any case.

Figure 3-2 on page 72 shows an example of the ASMI accessed from a Web browser.

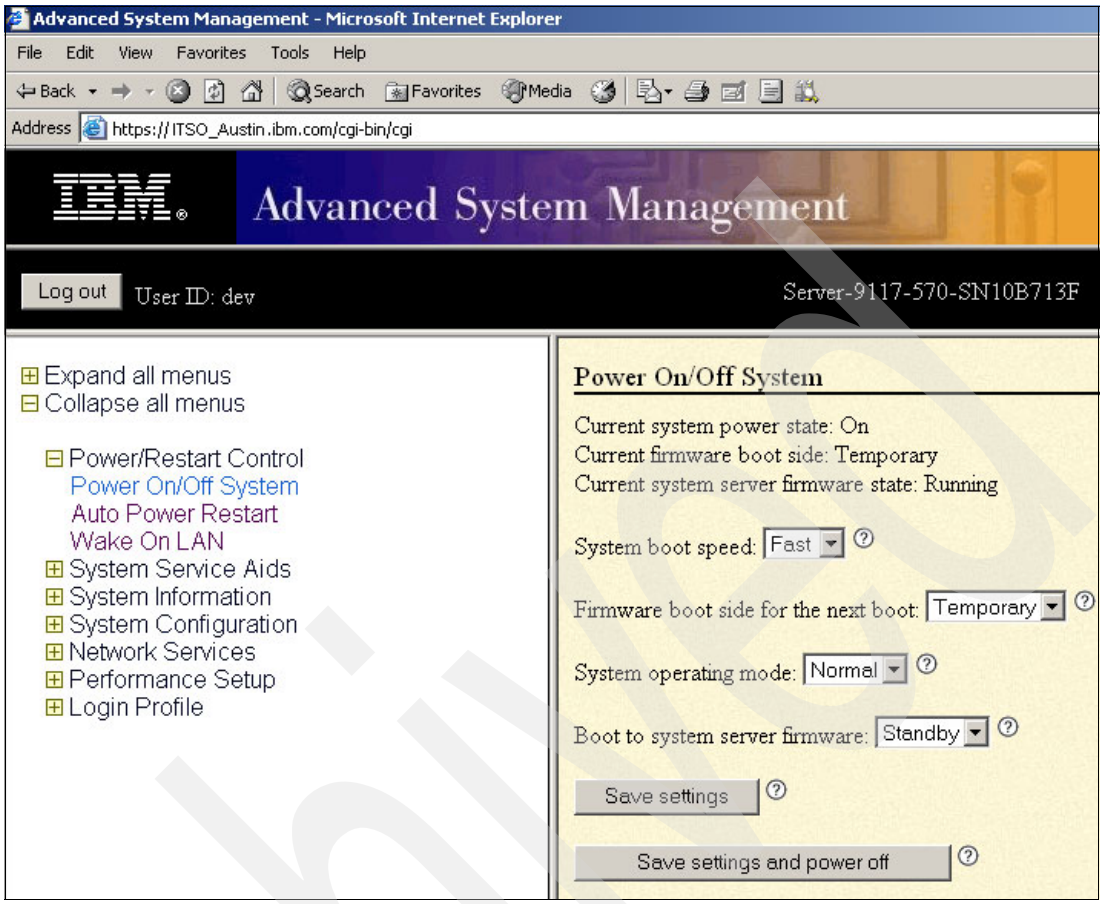


Figure 3-2 Advanced System Management main menu

3.2.2 Partition diagnostics

The diagnostics consist of stand-alone diagnostics, which are loaded from the DVD-ROM drive, and online diagnostics (available in AIX 5L):

- Online diagnostics, when installed, are resident with AIX 5L on the disk or server. They can be booted in single-user mode (service mode), run in maintenance mode, or run concurrently (concurrent mode) with other applications. They have access to the AIX 5L error log and the AIX 5L configuration data:
  - Service mode (requires service mode boot) enables the checking of system devices and features. Service mode provides the most complete checkout of the system resources. All system resources, except the SCSI adapter and the disk drives used for paging, can be tested.
  - Concurrent mode enables the normal system functions to continue while selected resources are being checked. Because the system is running in normal operation, some devices might require additional actions by the user or diagnostic application before testing can be done.
  - Maintenance mode enables the checking of most system resources. Maintenance mode provides the exact same test coverage as Service Mode. The difference between the two modes is the way they are invoked. Maintenance mode requires that all activity on the operating system is stopped. The **`shutdown -m`** command is used to stop all activity on the operating system and put the operating system into maintenance mode.

- The System Management Services (SMS) error log is accessible from the SMS menu for tests performed through SMS programs. For results of service processor tests, access the error log from the service processor menu.

Note: Because the p5-505 and p5-505Q have an optional DVD-ROM (FC 1903) and DVD-RAM (FC 1900), alternate methods for maintaining and servicing the system need to be available if the DVD-ROM or DVD-RAM is not ordered. It is possible to use Network Install Manager (NIM) server for this purpose.

3.2.3 Service Agent

Service Agent is an application program that operates on an IBM System p computer and monitors the computer for hardware errors. It reports detected errors, assuming they meet certain criteria for severity, to IBM for service with no intervention. It is an enhanced version of Service Director™ with a graphical user interface.

Key things you can accomplish using Service Agent for the IBM System p5, pSeries, and RS/6000 include:

- Automatic VPD collection
- Automatic problem analysis
- Problem-definable threshold levels for error reporting
- Automatic problem reporting where service calls are placed to IBM without intervention
- Automatic client notification

In addition, there are:

- Commonly viewed hardware errors. You can view hardware event logs for any monitored machine in the network from any Service Agent host user interface.
- High-availability cluster multiprocessing (HACMP) support for full fallback.
- Network environment support with minimum telephone lines for modems.
- Communication base provided for performance data collection and reporting tool Performance Management (PM/AIX). For more information about PM/AIX, see:

  `http://www.ibm.com/servers/aix/pmaix.html`

You define machines by using the Service Agent user interface. After you define the machines, they are registered with the IBM Service Agent Server (SAS). During the registration process, the registration process creates an electronic key that becomes part of your resident Service Agent program. You use this key each time the Service Agent places a call for service. The IBM Service Agent Server checks the current client service status from the IBM entitlement database. If this reveals that you are not on Warranty or MA, the IBM Service Agent Server refuses the service call and posts your call or a reply back using an e-mail notification.

You can configure Service Agent to connect to IBM either using a modem or a network connection. In any case, the communication is encrypted and strong authentication is used. Service Agent sends outbound transmissions only and does not allow any inbound connection attempts. Only hardware machine configuration, machine status, or error information is transmitted. Service Agent does not access or transmit any other data on the monitored systems.

Thee principal ways of communication are possible:

- Dial-up using attached modem device (uses the AT&T Global Network dialer for modem access, and it does not accept incoming calls to modem)
- VPN (IPsec is used in this case)
- HTTPS (can be configured to work with firewalls and authenticating proxies)

Figure 3-3 shows possible communication paths for an IBM System p5 system that is configured to use all the features of Service Agent. In this figure, communication to IBM support can be through either a modem or the network. If an HMC is present, Service Agent is an integral part of the HMC and, if activated, collects hardware-related information and error messages about the entire system and partitions. If software level information (such as performance data) is also required, Service Agent can also be installed on any of the partitions and can be configured to act as either a gateway and a connection manager or as a client. When Service Agent is configured as a gateway and a connection manager, it gathers data from clients and communicates to IBM on behalf of them.

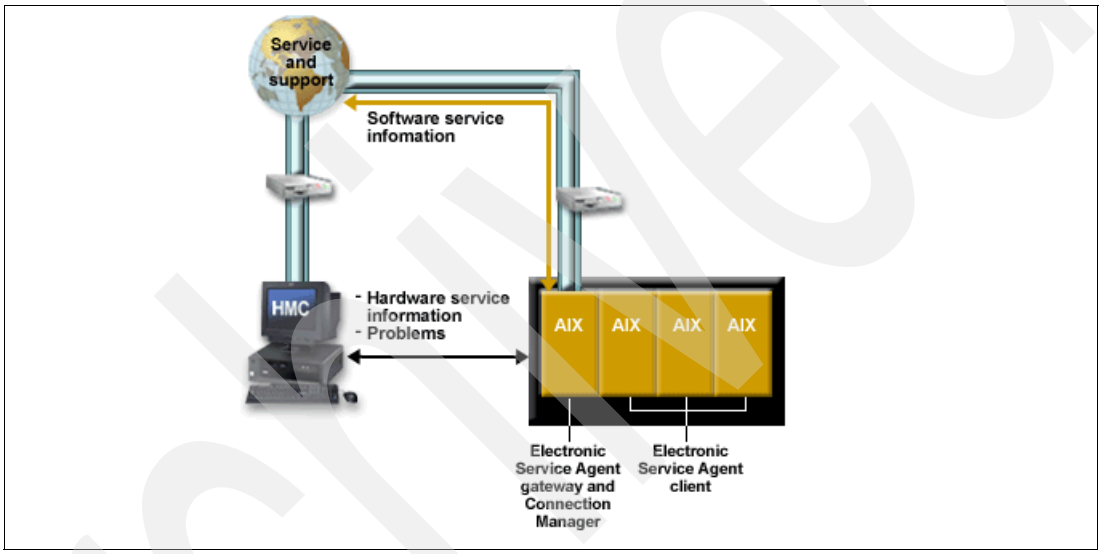


Figure 3-3 Service agent and possible connections to IBM

Additional services provided by Service Agent are:

- My Systems: The client and IBM employees authorized by the client can view hardware and software information and error messages that are gathered by Service Agent on Electronic Services WWW pages at:

  `http://www.ibm.com/support/electronic`

- Premium Search: Search service using information gathered by Service Agents (paid service that requires special contract).
- Performance Management: Service Agent provides a means for collecting long term performance data. The data is collected in reports accessed by the client on WWW pages of Electronic Services (paid service that requires special contract).

You can download the latest version of Service Agent at:

`ftp://ftp.software.ibm.com/aix/service_agent_code`

Service Focal Point

Traditional service strategies become more complicated in a partitioned environment. Each logical partition reports errors it detects, without determining if other logical partitions also detect and report the errors. For example, if one logical partition reports an error for a shared

resource, such as a managed system power supply, other active logical partitions might report the same error. The Service Focal Point application helps you to avoid long lists of repetitive call-home information by recognizing that these are repeated errors and correlating them into one error.

Service Focal Point is an application on the HMC that enables you to diagnose and repair problems on the system. In addition, you can use Service Focal Point to initiate service functions on systems and logical partitions that are not associated with a particular problem. You can configure the HMC to use the Service Agent call-home feature to send IBM event information. Service Focal Point is available also in Integrated Virtualization Manager. It allows you to manage serviceable events, create serviceable events, manage dumps, and collect vital product data (VPD) when no reporting via Service Agent is possible.

3.2.4 IBM System p5 firmware maintenance

The IBM System p5, pSeries, and RS/6000 Client-Managed Microcode is a methodology that enables you to manage and install microcode updates on IBM System p5, pSeries, RS/6000 systems, and associated I/O adapters. You can install the IBM System p5 microcode either from an HMC or from a running partition. For update details, see 2.14.6, “System firmware” on page 59.

If you use an HMC to manage your server, you can use the HMC interface to view the levels of server firmware and power subsystem firmware that are installed on your server and are available to download and install.

Each IBM System p5 server has the following levels of server firmware and power subsystem firmware:

- Installed level – This is the level of server firmware or power subsystem firmware that has been installed and will be installed into memory after the managed system is powered off and powered on. It is installed on the “*t*” side of system firmware. For an additional discussion about firmware sides, see 2.14.7, “Service processor” on page 61.
- Activated level – This is the level of server firmware or power subsystem firmware that is active and running in memory.
- Accepted level – This is the backup level of server or power subsystem firmware. You can return to this level of server or power subsystem firmware if you decide to remove the installed level. It is installed on the “*p*” side of system firmware. For an additional discussion about firmware sides, see 2.14.7, “Service processor” on page 61.

IBM introduced the Concurrent Firmware Maintenance (CFM) function on System p5 systems in system firmware level 01SF230_126_120, which was released on 16 June 2005. This function supports nondisruptive system firmware service packs to be applied to the system concurrently (without requiring a reboot to activate changes). For systems that are not managed by an HMC, the installation of system firmware is always disruptive.

The concurrent levels of system firmware can, on occasion, contain fixes that are known as deferred. These deferred fixes, which can be installed concurrently, are not activated until the next IPL. For deferred fixes within a service pack, only the fixes in the service pack, which cannot be concurrently activated, are deferred. Figure 3-4 on page 76 shows the system firmware file naming convention.

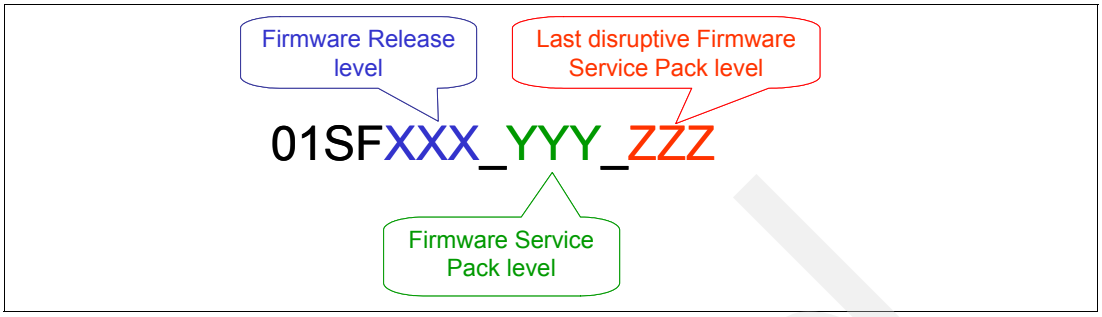


Figure 3-4 System firmware file naming convention

An installation is disruptive if:

- The release levels (*XXX*) of currently installed and new firmware are different.
- The service pack level (*YYY*) and the last disruptive service pack level (*ZZZ*) are equal in the new firmware.

Otherwise, an installation is concurrent if:

- The service pack level (*YYY*) of the new firmware is higher than the service pack level currently installed on the system and the above conditions for disruptive installation are not met.

3.3 Cluster solution

Today's IT infrastructure requires that servers meet increasing demands, while offering the flexibility and manageability to rapidly develop and deploy new services. IBM clustering hardware and software provide the building blocks, with availability, scalability, security, and single-point-of-management control, to satisfy these needs. The advantages of clusters are:

- Large-capacity data and transaction volumes, including support of mixed workloads
- Scale-up (add processors) or scale-out (add servers) without down time
- Single point-of-control for distributed and clustered server management
- Simplified use of IT resources
- Designed for 24x7 access to data applications
- Business continuity in the event of a disaster

The POWER processor-based AIX 5L and Linux cluster targets scientific and technical computing, large-scale databases, and workload consolidation. IBM Cluster Systems Management software (CSM) is designed to provide a robust, powerful, and centralized way to manage a large number of POWER5 processor-based servers, all from one single point-of-control. Cluster Systems Management can help lower the overall cost of IT ownership by helping to simplify the tasks of installing, operating, and maintaining clusters of servers. Cluster Systems Management can provide one consistent interface for managing both AIX 5L and Linux nodes (physical systems or logical partitions), with capabilities for remote parallel network install, remote hardware control, and distributed command execution.

Cluster Systems Management for AIX 5L and Linux on POWER processor-based servers is supported on the p5-505 and p5-505Q. For hardware control, an HMC is required. One HMC can also control several IBM System p5 servers that are part of the cluster. If a server that is configured in partition mode (with physical or virtual resources) is part of the cluster, all partitions must be part of the cluster.

Monitoring is much easier to use, and the system administrator can monitor all of the network interfaces, not just the switch and administrative interfaces. The management server pushes information out to the nodes, which releases the management server from having to trust the node. In addition, the nodes do not have to be network-connected to each other. This means that giving root access on one node does not mean giving root access on all nodes. The base security setup is all done automatically at install time.

For information regarding the IBM Cluster Systems Management for AIX 5L, HMC control, cluster building block servers, and cluster software available, visit the following links:

- Cluster 1600

  http://www.ibm.com/servers/eserver/clusters/hardware/1600.html

- Cluster 1350™

  http://www.ibm.com/systems/clusters/hardware/1350.html

The CSM ships with AIX 5L itself (a 60-day Try and Buy license is shipped with AIX). The CSM client side is installed automatically and is ready when you install AIX 5L. So, each system or logical partition is cluster-ready.

The CSM V1.5 on AIX 5L and Linux introduces an optional IBM CSM High Availability Management Server feature, which is designed to allow automated failover of the CSM management server to a backup management server. In addition, sample scripts for setting up Network Time Protocol (NTP), and network tuning (AIX 5L only) configurations, and the capability to copy files across nodes or node groups in the cluster can improve cluster ease of use and site customization.

A

Servicing an IBM System p5 system

POWER5 processor-based servers can be designated as one of the following types:

- Client setup (CSU) with client-installable features (CIF) and client-replaceable units (CRU)

  The p5-505 is considered CSU.
- Authorized service representative set up, upgraded, and maintained

A number of Web-based resources are available to assist clients and service providers in planning, installing, and maintaining servers.

Note: This applies to IBM System p5 and IBM eServer p5 in general.

Resource Link

Resource Link™ is a customized, Web-based solution, providing access to information for planning, installing, and maintaining IBM servers and associated software. It includes similar information about other selected IBM servers. Access to the site is by an IBM registration ID and password that are available free of charge. Resource Link pages can vary by user authorization level and are continually updated; therefore, the details that you see when accessing Resource Link might not exactly match what we mention here.

Resource Link contains links to:

- Education
- Planning
- Forums
- Fixes

Resource Link is available at:

http://www.ibm.com/servers/resourcelink

IBM Systems Hardware Information Center

The IBM Systems Hardware Information Center is a source for both hardware and software technical information for systems. It has information to help perform a variety of tasks, including:

- Preparing a site to accommodate the hardware for IBM systems.
- Installing the server, console, features and options, and other hardware.
- Installing and using a Hardware Management Console.
- Partitioning the server and installing the operating systems.
- Enabling and managing Capacity on Demand.
- Troubleshooting problems and servicing the server. Included here are component removal and replacement procedures and the Start of Call procedure.

  Physical components of a system are generally considered either a client-replaceable unit (CRU) or a field-replaceable unit (FRU). CRUs are further categorized as either Tier 1 CRUs or Tier 2 CRUs with the following definitions:

  - Tier 1 CRU: Very easy to replace
  - Tier 2 CRU: More complicated to replace
  - FRU: Replaced by the service provider

Removal and replacement procedures can be documented in the Information Center accompanied by graphics, such as Figure A-1 on page 81 and video clips.

Alternatively, they can take the form of guided procedures using the HMC, for example, by selecting **Service Applications** → **Service Focal Point** → **Exchange Parts**.

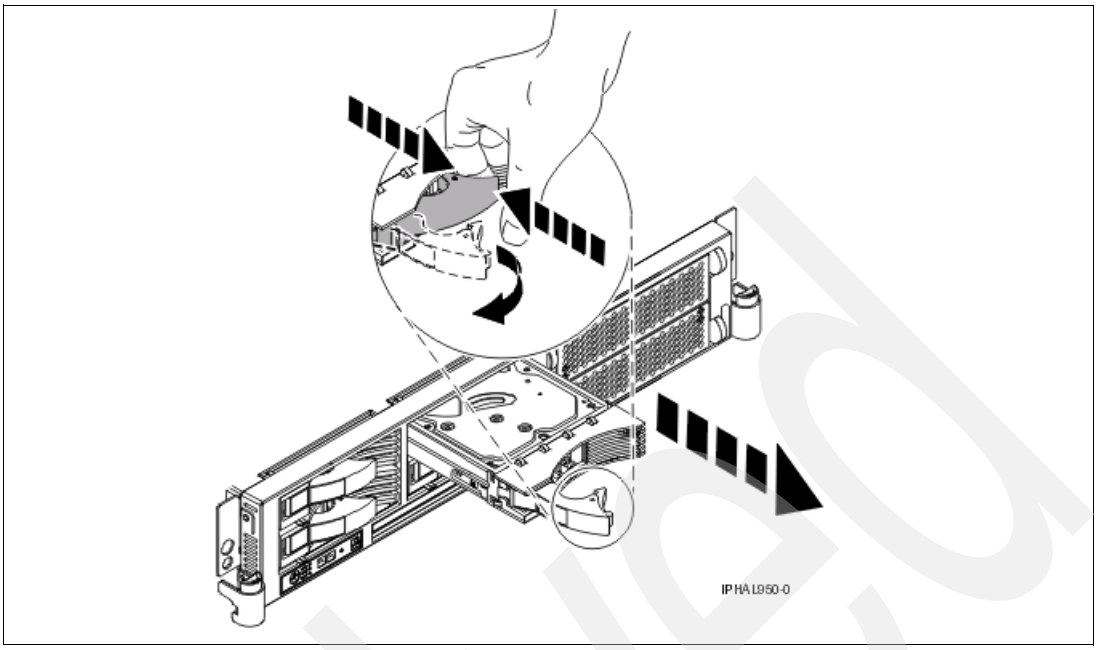


Figure A-1 Removing a disk drive

Note: Part classification, contractual agreements, and implementation in specific geographies all affect how CRUs and FRUs are determined.

The IBM Systems Hardware Information Center is available:

- On the Internet

  `http://www.ibm.com/servers/library/infocenter`
- On the HMC

  Click **Information Center and Setup Wizard** → **Launch the Information Center**.
- On CD-ROM:
  - Shipped with the hardware (English SK3T-8159)
  - Also available for you to order from IBM Publications Center

Related publications

The publications listed in this section are considered particularly suitable for a more detailed discussion of the topics covered in this Redpaper.

IBM Redbooks

For information about ordering these publications, see "How to get IBM Redbooks" on page 85. Note that some of the documents referenced here may be available in softcopy only.

- *Advanced POWER Virtualization on IBM System p5*, SG24-7940
- *Partitioning Implementations for IBM @server p5 Servers*, SG24-7039
- *Advanced POWER Virtualization on IBM @server p5 Servers: Architecture and Performance Considerations*, SG24-5768
- *IBM @server pSeries Sizing and Capacity Planning: A Practical Guide*, SG24-7071
- *IBM @server p5 590 and 595 System Handbook*, SG24-9119
- *LPAR Simplification Tools Handbook*, SG24-7231
- *Virtual I/O Server Integrated Virtualization Manager*, REDP-4061
- *IBM @server p5 590 and 595 Technical Overview and Introduction*, REDP-4024
- *IBM @server p5 510 Technical Overview and Introduction*, REDP-4001
- *IBM @server p5 520 Technical Overview and Introduction*, REDP-9111
- *IBM @server p5 550 Technical Overview and Introduction*, REDP-9113
- *IBM @server p5 570 Technical Overview and Introduction*, REDP-9117
- *IBM System p5 510 and 510Q Technical Overview and Introduction*, REDP-4136
- *IBM System p5 520 and 520Q Technical Overview and Introduction*, REDP-4137
- *IBM System p5 550 and 550Q Technical Overview and Introduction*, REDP-4138
- *IBM System p5 560Q Technical Overview and Introduction*, REDP-4139
- *Virtual I/O Server Integrated Virtualization Manager*, REDP-4061
- *Hardware Management Console (HMC) Case Configuration Study for LPAR Management*, REDP-3999

Other publications

These publications are also relevant as further information sources:

- *7014 Series Model T00 and T42 Rack Installation and Service Guide*, SA38-0577, contains information regarding the 7014 Model T00 and T42 Rack, in which this server can be installed.
- *7316-TF3 17-Inch Flat Panel Rack-Mounted Monitor and Keyboard Installation and Maintenance Guide*, SA38-0643, contains information regarding the 7316-TF3 Flat Panel Display, which can be installed in your rack to manage your system units.

- *IBM eServer Hardware Management Console for pSeries Installation and Operations Guide*, SA38-0590, provides information to operators and system administrators about how to use an IBM Hardware Management Console for pSeries (HMC) to manage a system. It also discusses the issues associated with logical partitioning planning and implementation.
- *Planning for Partitioned-System Operations*, SA38-0626, provides information to planners, system administrators, and operators about how to plan for installing and using a partitioned server. It also discusses issues associated with the planning and implementation of partitioning.
- *RS/6000 and eServer pSeries Diagnostics Information for Multiple Bus Systems*, SA38-0509, contains diagnostic information, service request numbers (SRNs), and failing function codes (FFCs).
- *System p5, eServer p5 Customer service support and troubleshooting*, SA38-0538, contains information regarding slot restrictions for adapters that can be used in this system.
- *System Unit Safety Information*, SA23-2652, contains translations of safety information used throughout the system documentation.

Online resources

These Web sites and URLs are also relevant as further information sources:

- AIX 5L operating system maintenance packages downloads

  http://www.ibm.com/servers/eserver/support/unixservers/aixfixes.html
- IBM Systems Hardware Information Center documentation

  http://publib.boulder.ibm.com/infocenter/eserver/v1r3s/index.jsp
- IBM Systems Information Centers

  http://publib.boulder.ibm.com/eserver/
- IBM microcode downloads

  http://www14.software.ibm.com/webapp/set2/firmware/gjsn
- Support for IBM System p servers

  http://www.ibm.com/servers/eserver/support/unixservers/index.html
- Technical help database for AIX 5L

  http://www14.software.ibm.com/webapp/set2/srchBroker/views/srchBroker.jsp?rs=111
- IBMlink

  http://www.ibmlink.ibm.com
- Linux on System p

  http://www.ibm.com/systems/p/linux/
- Microcode Discovery Service

  http://www14.software.ibm.com/webapp/set2/mds/fetch?page=mds.html

How to get IBM Redbooks

You can search for, view, or download Redbooks, Redpapers, Hints and Tips, draft publications and Additional materials, as well as order hardcopy Redbooks or CD-ROMs, at this Web site:

ibm.com/redbooks

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services

IBM System p5 505 and 505Q

Technical Overview and Introduction

Redpaper

High-performance server in a dense, 1U package is ideal for data centers with limited resources

New option for support Micro-Partitioning technology without an HMC to help lower TCA and TCO

The raw computing power for high-performance engineering and scientific workloads

This IBM Redpaper is a comprehensive guide covering the IBM System p5 505 server supporting the IBM AIX 5L and Linux operating systems. We introduce major hardware offerings and discuss their prominent functions.

Professionals wanting to acquire a better understanding of IBM System p5 products should consider reading this document. The intended audience includes:

- Clients
- Marketing representatives
- Technical support professionals
- IBM Business Partners
- Independent software vendors

This document expands the current set of IBM System p5 documentation by providing a desktop reference that offers a detailed technical description of the p5-505.

This publication does not replace the latest IBM System p5 marketing materials, tools, or product documentation. It is intended as an additional source of information that, together with existing sources, can be used to enhance your knowledge of IBM server solutions.

INTERNATIONAL TECHNICAL SUPPORT ORGANIZATION

BUILDING TECHNICAL INFORMATION BASED ON PRACTICAL EXPERIENCE

IBM Redbooks are developed by the IBM International Technical Support Organization. Experts from IBM, Customers and Partners from around the world create timely technical information based on realistic scenarios. Specific recommendations are provided to help you implement IT solutions more effectively in your environment.

For more information:
ibm.com/redbooks