Draft Document for Review February 27, 2006 8:01 pm

REDP-4137-00

IBM System p5 520 and 520Q Technical Overview and Introduction

- Finer system granulation using Micro-Partitioning technology to help lower TCO
- Support for versions of AIX 5L and Linux operating systems
- From Web servers to integrated cluster solutions

Giuliano Anselmi
Bernard Filhol
SahngShin Kim
Gregor Linzmeier
Ondrej Plachy

Redpaper

ibm.com/redbooks

International Technical Support Organization

IBM System p5 520 and 520Q Technical Overview and Introduction

March 2006

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

First Edition (March 2006)

This edition applies to IBM System p5 520 (product number 9131-52A), 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, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

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

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

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

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

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

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

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

Trademarks

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

@server®
@server®
Redbooks (logo)  ™
pSeries®
AIX 5L™
AIX®
Chipkill™
Enterprise Storage Server®
FICON®
HACMP™
IBM®
Micro-Partitioning™
OpenPower™
PowerPC®
POWER™
POWER4™
POWER5™
POWER5+™
Redbooks™
RS/6000®
Service Director™
System p™
System p5™
System Storage™
TotalStorage®
Virtualization Engine™

The following terms are trademarks of other companies:

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

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

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

Linux is a trademark of Liux 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 document is a comprehensive guide covering the IBM® System p5™ 520 and 520Q UNIX® servers. It introduces major hardware offerings and discusses their prominent functions.

Professionals wishing to acquire a better understanding of IBM System p™ products should read this document. The intended audience includes:

- Clients
- Sales and marketing professionals
- Technical support professionals
- IBM Business Partners
- Independent software vendors

This document expands the current set of IBM System p documentation by providing a desktop reference that offers a detailed technical description of the p5-520 and the p5-520Q system.

This publication does not replace the latest IBM System p marketing materials and tools. It is intended as an additional source of information that, together with existing sources, may 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.

Giuliano Anselmi is a certified pSeries® Presales Technical Support Specialist working in the Field Technical Sales Support group based in Rome, Italy. For seven years, he was an IBM @server pSeries Systems Product Engineer, supporting Web Server Sales Organization in EMEA, IBM Sales, IBM Business Partners, Technical Support Organizations, and IBM Dublin @server Manufacturing. Giuliano has worked for IBM for 14 years, devoting himself to RS/6000® and pSeries systems with his in-depth knowledge of the related hardware and solutions.

Bernard Filhol is a UNIX Server Customer Satisfaction Resolution Team Leader for NEE and SWE IOTs in Montpellier France. He has more than 25 years of experience in mainframes and five years of experience in pSeries Customer Satisfaction. He holds a degree in Electronic from Montpelier University Institute of Technology. His areas of expertise include Mainframe Channel Subsystem, FICON® and pSeries RAS. He has written extensively on FICON.

SahngShin Kim is a sales specialist of STG infra-solution sales team in Seoul, Korea. For three years he was a sales specialist of IBM @server pSeries and for two years of grid computing and one year for infra-solutions. SahngShin has worked for IBM for six years, devoting himself to RS/6000 and pSeries systems and STG server products and architecting of those.

Gregor Linzmeier is an IBM Advisory IT Specialist for RS/6000 and pSeries workstation and entry servers as part of the Systems and Technology Group in Mainz, Germany supporting

IBM sales, Business Partners, and clients with pre-sales consultation and implementation of client/server environments. He has worked for more than 15 years as an infrastructure specialist for RT, RS/6000, and AIX® in large CATIA client/server projects.

Ondrej Plachy is an IT specialist in IBM Czech Republic responsible for project design, implementation, and support of large scale computer systems. He has 11 years of experience in the UNIX field. He holds the Ing. academic degree in Computer Science from Czech Technical University (CVUT), Prague. He has worked at Supercomputing Centre of Czech Technical University for four years, and currently works in IBM for seven years in AIX 5L™ support team.

The project that produced this document was managed by:

Scott Vetter
IBM U.S.

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

Larry Amy, Baba Arimilli, Ron Arroyo, Terry Brennan, Erin Burke, Mark Dewalt, Bob Foster, Ron Gonzalez, Dan Henderson, Hal Jennings, Carolyn Jones, Bill Mihaltse, Thoi Nguyen, Ken Rozendal, Craig Shempert, Dave Willoughby, David A. Hepkin, Brian J King.
IBM U.S.

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. JN9B Building 905
  11501 Burnet Road
  Austin, Texas 78758-3493

1

General description

The IBM System p5 520 and 520Q rack-mount and deskside servers (9131-52A) give you new tools for managing on demand business, greater application flexibility, and innovative technology in 1-core, 2-core, and 4-core configurations—all designed to help you capitalize on the on demand business revolution. To simplify naming, both products are referred to as “p5-520” or p5-520Q” server.

The p5-520 and p5-520Q server have the POWER5+™ microprocessor that provides performance and reliability advances (or enhancements) over the POWER5™ architecture it replaces. Chief among the enhancements is 90 nm chip design technology.

The p5-520 processor is packaged as a 1-core Single-Core Module running at 1.65 GHz with no L3 cache or a 2-core Dual-Core Module running at 1.65 or 1.9 GHz with 36 MB of L3. The p5-520Q offers the same features but comes with a 4-core POWER5+ Quad-Core Module running at 1.5 GHz with two 36 MB of L3 caches.

When you purchase a System p5-520 or p5-520Q Express Edition that is only available on an initial order request, you may qualify for processor activation at no extra charge. The number of processors, total memory, quantity and size of disk, and the presence of media device are the only features that determine if you are entitled to a processor entitlement at no additional charge. Contact your sales representative regarding the feature for Express Edition or volume offering.

The p5-520 and p5-520Q server have a base of 1 GB of DDR2 memory that can be expanded to 32 GB designed for performance and exploitation of 64-bit addressing as used in large database applications.

The p5-520 and p5-520Q includes four front-accessible, hot-swap-capable disk bays in a minimum configuration with an additional four hot-swap-capable disk bays as an optional feature. The eight disk bays can accommodate up to 2.4 TB of disk storage using the 300 GB Ultra320 SCSI disk drives. Other features included in the p5-520 and p5-520Q are six hot-plug PCI-X slots with Enhanced Error Handling (EEH), integrated service processor, integrated 10/100/1000 Mbps two-port Ethernet, two system, two USB, and two HMC ports, integrated dual-channel Ultra320 SCSI controller, hot-swappable power and cooling, and optional redundant power.

Three non-hot-swap media bays are used to accommodate additional devices. Two media bays only accept slim line media devices, such as DVD-ROM or DVD-RAM drives, and one half-height bay is used for a tape drive. The rack-mount model also has I/O extension capability using the RIO-2 bus that allows attachment of the 7311 Model D20 I/O drawers.

For partitioning, a Hardware Management Console (HMC) is recommended. Dynamic LPAR is supported on the p5-520 and p5-520Q servers, allowing up to two logical partitions. In addition, the optional Advanced POWER™ Virtualization feature supports up to 40 micro-partitions using Micro-Partitioning™ technology. The Integrated Virtualization Manager provides partition management in settings where an HMC is unavailable or not desired.

Additional reliability and availability features include redundant hot-swap cooling fans and redundant power supplies. Along with these components, the p5-520 and p5-520Q are designed to provide an extensive set of reliability, availability, and serviceability (RAS) features that include improved with dual service processor; fault isolation, recovery from errors without stopping the system, avoidance of recurring failures, and predictive failure analysis.

The p5-520 and p5-520Q servers are backed by a three-year limited warranty. Check with your IBM representative for particular warranty availability in your region.

1.1 System specifications

Table 1-1 lists the general system specifications of the p5-520 and p5-520Q systems.

Table 1-1 IBM System p5 520 and p5 520Q specifications

Description	Range
Operating temperature	5 to 35 degrees Celsius (41 to 95 F)
Relative humidity	8% to 80%
Operating voltage	100 to 127 or 200 to 240 V ac (auto-ranging)
Operating frequency	47/63 Hz
Maximum power consumption	850 watts maximum
Maximum thermal output	2936.36 Btu/hour (maximum)

1.2 Physical package

The following sections discuss the major physical attributes found on the p5-520 and p5-520Q systems in rack-mounted and deskside versions selectable through a feature code.

1.2.1 Deskside model

The p5-520 or and p5-520Q can be configured as a deskside models. Figure 1-1 on page 4 shows the system and Table 1-2 provides a list of physical attributes.[1]

Table 1-2 Physical packaging of the deskside model

Dimension[a]	Deskside (FC 7919)
Height	533 mm (21.0 inches)
Width	203 mm (8.0 inches)
Depth (without rear cover; FC 6587)	584.0 mm (23.0 inches)
Depth (with rear cover; FC 6587)	706.0 mm (27.8 inches)
Weight	
Weight	43 kg (95 lb)
Shipping weight	50 kg (110 lb)

[a]For specific region, such as China, check specifications for specific dimensions.

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

Figure 1-1 The deskside model (FC 7184) and acoustic cover (right FC 7185)

The p5-520 or p5-520Q, when configured as a deskside server, is ideal for environments requiring the user to have local access to the machine. A typical example of this would be applications requiring a native graphics display.

To order a system as a deskside version, FC 7184 or FC 7185 is required. FC 7185 is designed for quiet operation in office environments. The system is designed to be set up by the client and, in most cases, will not require the use of any tools. Full set-up instructions are included with the system.

The GXT135P 2D graphics accelerator with analog and digital interfaces (FC 1980) is available and is supported for SMS, firmware menus, and other low-level functions, as well as when AIX 5L or Linux® starts the X11-based graphical user interface. You can use graphical AIX 5L system tools for configuration management if the adapter is connected to the primary console, such as the IBM 15-inch, 17-inch, 19-inch, 20-inch TFT Color Monitor (FC 3641, FC 3645, FC 3644, and FC 3643).

1.2.2 Rack-mount model

IBM System p5-520 or p5-520Q can be configured as a 4U rack-mount model with the selected feature code. Figure 1-2 on page 5 shows the system and Table 1-3 provides a list of physical attributes.

Table 1-3 Physical packaging of the rack-mount model

Dimension[a]	**Rack (FC 7918)**
Height	178 mm (7.0 inches)
Width	437 mm (17.2 inches)
Depth	584 mm (23.0 inches)
Weight	
Weight	43.0 kg (95 lb)
Shipping weight	53.0 kg (117 lb)

[a] For specific region, such as China, check specifications for specific dimensions.

Figure 1-2 IBM System p5-520 and p5-520Q rack-type model (FC 7160)

The p5-520 or p5-520Q, when configured as a 4U rack-mounted server, is intended to be installed in a 19-inch rack, thereby enabling efficient use of computer room floor space. If the IBM 7014 T42 rack is used to mount the server, it is possible to place up to 10 systems in an area of 644 mm (25.5 inches) x 1147 mm (45.2 inches).

To order a p5-520 or p5-510Q system as a rack-mounted version, FC 7190 must be selected. In addition to the rack-mounted version, the server can be installed in either IBM or OEM racks. Therefore, you are required to select one of the following features:

- IBM Rack-mount Drawer Rail Kit (FC 7160)
- OEM Rack-mount Drawer Rail Kit (FC 7161)

Included with the rack-mounted server packaging are all of the components and instructions necessary to enable installation in a 19-inch rack using suitable tools.

The GXT135P 2D graphics accelerator with analog and digital interfaces (FC 1980) is available and is supported for SMS, firmware menus, and other low-level functions, as well as when AIX 5L or Linux starts the X11-based graphical user interface. You can use graphical AIX 5L system tools for configuration management if the adapter is connected to a common maintenance console, such as the 7316-TF3 rack-mounted flat-panel display.

1.3 Minimum and optional features

The systems are based on a flexible, modular design based on POWER5+ processors. The server is available in 1-core, 2-core, and 4-core configuration featuring:

- 1.65 GHz (SCM and DCM), 1.9 GHz (DCM), and 1.5 GHz (QCM) POWER5+ processors
- From 1 GB to 32 GB of total system memory capacity using 533 MHz DDR2 DIMM technology
- Four SCSI disk drives in a minimum configuration, eight SCSI disk drives with an optional second 4-pack enclosure for a total internal storage capacity of 2.4 TB using 300 GB disk drives
- Six PCI-X slots (one 266 MHz 64-bit PCIX-2, three 133 MHz 64-bit PCI-X, two 66 MHz 32-bit PCI-X). All slots support Enhanced Error Handling (EEH).

- Two slim-line media bays for optional storage devices
- One half-high bay for an optional tape device

The p5-520 and p5-520Q, including the service processor described in 3.2.1, “Service processor” on page 77, supports the following native ports:

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

  Optionally, an external USB diskette drive 1.44 (FC 2591) is available.
- Two HMC ports
- Optional GX+ Bus to RIO-2 adapter card (FC 2888)
- Two SPCN ports

In addition, the p5-520 and p5-520Q feature one internal Ultra320 SCSI dual channel controller, redundant hot-swap power supply (optional), and cooling fans.

The system supports 32-bit and 64-bit applications and requires a specific levels of AIX 5L and Linux operating systems. See 2.14, “Operating system support” on page 57.

1.3.1 Processor features

The p5-520 server features one or two POWER5+ chips with one, and two processor cores running at 1.65 GHz and 1.9 GHz or the p5-520Q with four 1.5 GHz cores. Processors are installed on either Single-Core Module (SCM), Dual-Core Module (DCM) or Quad-Core Module (QCM). For a list of available features, see Table 1-4.

Table 1-4 Processor feature codes

Feature code	Description
8321	1-core 1.65 GHz POWER5+ Processor Card, no L3 Cache
8323	2-core 1.65 GHz POWER5+ Processor Card, 36 MB L3 Cache
8330	2-core 1.9 GHz POWER5+ Processor Card, 36 MB L3 Cache
8333	4-core 1.5 GHz POWER5+ Processor Card, 2 x 36 MB L3 Cache

The POWER5+ chips are directly mounted to the system planar.

1.3.2 Memory features

The minimum memory requirement for the p5-520 and p5-520Q servers is 1 GB, and the maximum capacity is 32 GB. 533 MHz DDR2 technology. The planar of each system has eight sockets for memory DIMMs. Note that an amount of memory is always in use by the Hypervisor, even when the machine is not partitioned. The LPAR validation 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 1-5 on page 7 lists the available memory features.

Table 1-5 Memory feature codes

Feature code	Description
1930	1 GB (2 x 512 MB) DIMMs, 276-pin DDR2, 533 MHz SDRAM
1931	2 GB (2 x 1 GB) DIMMs, 276-pin DDR2, 533 MHz SDRAM
1932	4 GB (2 x 2 GB) DIMMs, 276-pin DDR2, 533 MHz SDRAM
1934	8 GB (2 x 4 GB) DIMMs, 276-pin DDR2, 533 MHz SDRAM

1.3.3 Disk and media features

The minimum configuration includes a 4-pack disk drive enclosure. A second 4-pack disk drive enclosure can be installed by ordering FC 6574 or FC 6594 so that the maximum internal storage capacity reach 2.4 TB (using the disk drive features available at the time of writing). The p5-520 and p5-520Q feature up to eight disk drive bays, two slim-line media device bays, and one half-height media bay. The minimum configuration requires at least one disk drive. Table 1-6 shows the disk drive feature codes that each bay can contain.

Table 1-6 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

Any combination of DVD-ROM and DVD-RAM drives of the following devices can be installed in the two slim-line bays:

- DVD-RAM drive, FC 1993
- DVD-ROM drive, FC 1994

A logical partition running a supported release of Linux requires a DVD-ROM drive or DVD-RAM drive to provide a way to run the diagnostics CD for hardware diagnostics. Concurrent diagnostics, as provided by the AIX 5L **diag** command, is not available on the Linux operating system at the time of writing.

Supplementary devices can be installed in the half-height media bay, such as:

- Internal 4 mm 36/72 GB LVD tape drive, FC 1991
- IBM 80/160 GB internal tape drive VXA, FC 1992
- IBM 160/320 GB internal tape drive with VXA-3 technology, FC 1892
- IBM 200/400 GB LTO2 tape drive, FC 1997

Devices installed in the media bays must be assigned as a group to a single LPAR on a partitioned system.

A dual-channel RAID enablement daughter card is also available (FC 1907).

1.3.4 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 inch) 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.

1.3.5 I/O drawers

The p5-520 and p5-520Q have six internal PCI-X slots, where three of them are long slots and three are short slots. If more PCI-X slots are needed, especially well-suited to extend the number of LPARs and partitions, up to four 7311 Model D20 drawers can be connected to the two RIO-2 ports on the rear of the system that are provided in a minimum configuration.

The 7311 Model D20 I/O drawer is a 4U full-size drawer, which must be mounted in a rack. It features seven hot-pluggable PCI-X slots and optionally up to 12 hot-swappable disks arranged in two 6-packs. Redundant, concurrently maintainable power and cooling is an optional feature (FC 6268). The 7311 Model D20 I/O drawer offers a modular growth path for a system with increasing I/O requirements. When a p5-520 or p5-520Q is fully configured with four attached 7311 Model D20 drawers, the combined system supports up to 34 PCI-X adapters (in a maximum configuration (Remote I/O expansion cards are required) and 56 hot-swappable SCSI disks, for a total internal capacity of 16.8 TB using 300 GB disks.

PCI-X and PCI cards are inserted from the top of the I/O drawer down into the slot from the drawers front service position. The installed adapters are protected by plastic separators, designed to prevent grounding and damage when adding or removing adapters.

The drawer has the following attributes:

- 4U rack-mount enclosure assembly
- Seven PCI-X slots 3.3 volt, keyed, 133 MHz hot-pluggable
- Two 6-pack hot-swappable SCSI bays (optional)
- Optional redundant hot-swap power
- Two RIO-2 ports and two SPCN ports

Note: A 7311 Model D20 I/O drawer initial order, or an existing 7311 Model D20 I/O drawer that is migrated from another pSeries system, must have the RIO-2 ports available (FC 6417).

The I/O drawer has the following physical characteristics:

- Width: 482 mm (19.0 inches)
- Depth: 610 mm (24.0 inches)
- Height: 178 mm (7.0 inches)
- Weight: 45.9 kg (101 pounds)

Figure 1-3 on page 9 shows the different views of the 7311-D20 I/O drawer.

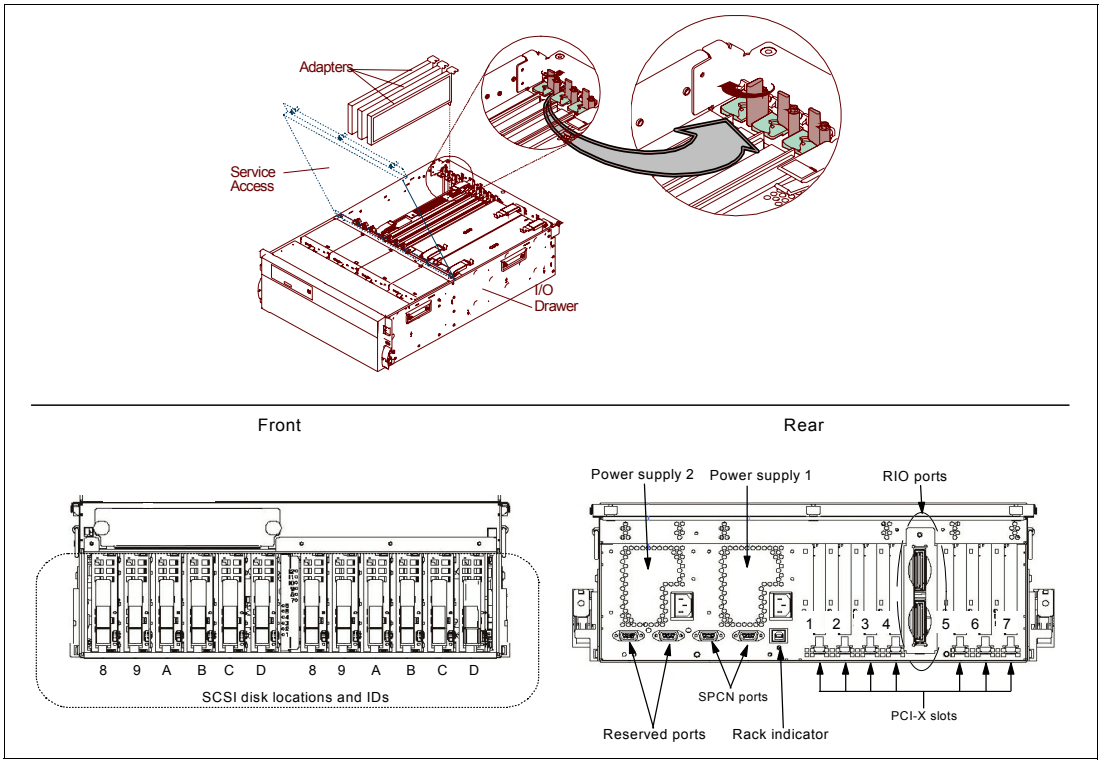


Figure 1-3 7311-D20 I/O drawer views

Note: The 7311 Model D20 I/O drawer is designed to be installed by an IBM service representative. Only the 7311 Model D20 I/O drawer is supported on a p5-520 or p5-520Q system.

1.3.6 Hardware Management Console models

A p5-520 or p5-520Q can be either HMC managed or non HMC managed. In HMC managed mode, an HMC (Hardware Management Console) 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 allocated to an individual partition.

Note: Non HMC managed modes are:

- Full system partition mode (only one partition containing all system resources exists on the system)
- Using the Integrated Virtualization Manager (IVM), see *2.12.5, "Integrated Virtualization Manager" on page 52*

See 2.13, "Hardware Management Console" on page 54 for detailed information about the HMC.

Table 1-7 on page 10 lists the HMC options for POWER5 processor-based systems available at the time of writing. Existing HMC models can be also used.

Table 1-7 Supported HMC

Type-model	Description
7310-C04	IBM 7310 Model C04 Desktop Hardware Management Console
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 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 C04 is a desktop model and the 7310 Model C05 is a deskside model with only 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.

In HMC managed installations with very high demand for high availability deployment of two HMCs should be considered. The service processor allows for connection of two HMCs, there is no need for special handling of such dual HMC environment. HMCs provide locking mechanism so that only one HMC at a time have write access to service processor.

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

Note: It is not possible to connect POWER4™ with POWER5 or 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.

1.4 Express Edition product offerings

The Express Edition configurations provide a convenient way to order any of several configurations 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.

1.4.1 Express Editions

If you order a System p5-520 server with a 1-core (FC 8321) or 2-core (FC 8323) POWER5+ 1.65 GHz processor, or a System p5-520Q server with a 4-core (FC 8333) 1.5 GHz processor Express Product Offerings as defined here, you may qualify for half of processor activation at no additional charge.

When an Express Edition is ordered, the configurator offers a choice of starting points that can be added onto. Clients can configure systems with one or two processor cards and two or four processor activations.

With the purchase of an Express Edition, 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 2 GB of memory installed for each active processor

If you order a System p5-520 server Express Edition as defined here, you may qualify for a processor activation at no extra charge. The number of processors, total memory, quantity/size of disk, and presence of a media device are the only features that determine if a client is entitled to a processor entitlement at no additional charge.

When you purchase an Express Edition, you will be entitled to a lower priced AIX 5L or Linux operating system license, or may choose to purchase the system with no operating system. The lower priced AIX 5L or Linux operating system is processed via a feature number on AIX 5L and either Red Hat or SUSE Linux. You may 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 will 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, AIX 5L edition; and systems with a lower priced Linux operating system will be referred to as the IBM System p5 Express, OpenPower™ edition.

In the case of Linux, only the fist subscription purchased is lower priced so, for example, additional licenses purchased for Red Hat to run in multiple partitions will be at 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.

If the system was initially ordered as an Express Edition, the system can be expanded at a later time using Express Edition pricing, when additional processors and activations along with the required memory are ordered on the same hardware upgrade order. The upgraded p5-520Q configuration must satisfy the Express Edition requirements for disk drives, memory, and processors. However, if the selection of total memory or disk drives are smaller than the total defined as the minimums, it disqualifies the order as an Express Product Offering.

1.4.2 Configurator starting points for Express Edition

The complete order must include the features identified in the minimum configuration plus the feature enhancements listed in Table 1-8.

Table 1-8 Express Edition feature code combinations

Offering				
Product Offering ID	**91311K1**	**91311K2**	**91311K3**	**91311K4**
Express Edition feature codes	9564	9565	9566	9567
1-core 1.65 GHz processor card	FC 8321 x 1	FC 8321 x 1	n/a	
2-core 1.65 GHz processor card	n/a	n/a	FC 8323 x 1	FC 8323 x 1
2-core 1.9 GHz processor card	n/a	n/a	n/a	
4-core 1.5 GHz processor card	n/a	n/a	n/a	n/a
1024 MB (2 x 512 MB) Memory DIMMs	FC 1930 x 1	FC 1930 x 1	FC 1930 x 2	FC 1930 x 2
2048 MB (2 x 1024 MB) Memory DIMM	n/a	n/a	n/a	n/a
73.4 GB 10,000 rpm disk drive	FC 1968 x 2	FC 1968 x 2	FC 1968 x 2	FC 1968 x 2
IBM deskside bezel and hardware	FC 7919 x 1	n/a	FC 7919 x 1	n/a
IBM rack-mount drawer bezel	n/a	FC 7190 x 1	n/a	FC 7190 x 1
Rack-mount drawer rail kit	n/a	FC 7160 x 1	n/a	FC 7160 x 1
Power Supply, 850 Watt	FC 5159 x 1	FC 5159 x 1	FC 5159 x 1	FC 5159 x 1

Offering				
Product Offering ID	**91311K1**	**91311K2**	**91311K3**	**91311K4**
Slimline IDE DVD-ROM	FC 1994 x 1	FC 1994 x 1	FC 1994 x 1	FC 1994 x 1
Media backplane	FC 7877 x 1	FC 7877 x 1	FC 7877 x 1	FC 7877 x 1
4-pack disk drive enclosure	FC 6574 x 1	FC 6574 x 1	FC 6574 x 1	FC 6574 x 1
Processor Entitlement	FC 7308 x 1	FC 7308 x 1	FC 7309 x 1	FC 7309 x 1
Product Offering ID	**91311K5**	**91311K6**	**91311K7**	**91311K8**
Express Edition feature codes	9568	9569	9571	9572
1-core 1.65 GHz processor card	n/a	n/a	n/a	n/a
2-core 1.65 GHz processor card	n/a	n/a	n/a	n/a
2-core 1.9 GHz processor card	FC 8330 x 1	FC 8330 x 1	n/a	n/a
4-core 1.5 GHz processor card	n/a	n/a	FC 8333 x 1	FC 8333 x 1
1024 MB (2 x 512 MB) Memory DIMMs	FC 1930 x 2	FC 1930 x 1	n/a	n/a
2048 MB (2 x 1024 MB) Memory DIMM	n/a	n/a	FC 1931 x 4	FC 1931 x 4
73.4 GB 10,000 rpm disk drive	FC 1968 x 2	FC 1968 x 1	FC 1968 x 2	FC 1968 x 2
IBM deskside bezel and hardware	FC 7919 x 1	n/a	FC 7919 x 1	n/a
IBM rack-mount drawer bezel	n/a	FC 7190 x 1	n/a	FC 7190 x 1
Rack-mount drawer rail kit	n/a	FC 7160 x 1	n/a	FC 7160 x 1
Power Supply, 850 Watt	FC 5159 x 1	FC 5159 x 1	FC 5159 x 1	FC 5159 x 1
Slimline IDE DVD-ROM	FC 1994 x 1	FC 1994 x 1	FC 1994 x 1	FC 1994 x 1
Media Backplane	FC 7877 x 1	FC 7877 x 1	FC 7877 x 1	FC 7877 x 1
4-pack disk drive Enclosure	FC 6574 x 1	FC 6574 x 1	FC 6574 x 1	FC 6574 x 1
Processor Entitlement	FC 7320 x 1	FC 7320 x 1	FC 7337 x 2	FC 7337 x 2

1.5 System racks

The IBM 7014 Model S11, S25, T00, and T42 Racks are 19-inch racks for general use with IBM System p5, IBM @server® p5, pSeries, and OpenPower Edition rack-mount servers. The racks provide increased capacity, greater flexibility, and improved floor space utilization.

If a server is to be installed in a non-IBM rack or cabinet, you must ensure that the rack conforms to the EIA[2] standard EIA-310-D (see 1.5.9, “OEM rack” on page 20).

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.

[2] 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.

1.5.1 IBM 7014 Model T00 Rack

The 1.8-meter (71-inch) Model T00 is compatible with past and present IBM System p systems. 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 1-4 on page 17), but others can fit inside the rack. See 1.5.6, “The ac power distribution unit and rack content” on page 16.
- An optional rack status beacon (FC 4690). This beacon is designed to be placed on top of a rack and cabled to servers, and other components inside the rack. Servers can be programmed to illuminate the beacon in response to a detected problem or changes in system status.
- A rack status beacon junction box (FC 4693) should be used to connect multiple servers to the beacon. This feature provides six input connectors and one output connector for the rack. To connect the servers or other components to the junction box or the junction box to the rack, status beacon cables (FC 4691) are necessary. Multiple junction boxes can be linked together in a series using daisy chain cables (FC 4692).
- Weights:
  - T00 base empty rack: 244 kg (535 pounds)
  - T00 full rack: 816 kg (1795 pounds)

1.5.2 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 client to more densely populate individual racks freeing valuable floor space without the need to purchase additional air conditioning units. The Rear Door Heat eXchanger features:

- Water-cooled heat exchanger door 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:

- Approximate height: 1945.5 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:

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

1.5.3 IBM 7014 Model S11 Rack

The Model S11 rack will 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; and 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).

1.5.4 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 so that it 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 inches) with side panels
- Depth: 1001 mm (39.4 inches) with front door
- Height: 1344 mm (49.0 inches)
- 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).

1.5.5 S11 rack and S25 rack considerations

The S11 and S25 racks do not have vertical mounting space that will accommodate FC 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 Customer Specified Rack Placement provides 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 1-9 on page 16 lists the machine types supported in the S11 and S25 racks.

Table 1-9 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	EXP4 Disk Enclosure	X	X
7311-D20	I/O Expansion Drawer	X	X
9110-510	@server p5 510	X	X
9111-520	@server p5-520	X	X
9113-550	@server p5 550	X	X
9115-505	System p5 505	X	X
9123-710	OpenPower 710	X	X
9124-720	OpenPower 720	X	X
9110-51A	System p5 510 and 510Q	X	X
9131-52A	System p5-520 and 520Q	X	X
9133-55A	System p5 550 and 550Q	X	X
9116-561	System p5 560Q	X	X
9910-P33	3000VA UPS (2700 watt)	X	X
9910-P65	500VA UPS (208-240V)		X
7315-CR3	Rack-mount HMC		X
7310-CR3	Rack-mount HMC		X
7026-P16	LAN attached async. RAN		X
7316-TF3	Rack-mounted flat-panel console kit		X

1.5.6 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, which are not included in the base order. For maximum availability it is highly recommended to connect power cords from 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. See Figure 1-4 on page 17 for 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 will be 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 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 the publication *IBM System p5, @server p5 and i5, and OpenPower Edition 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/eserver/

Note: Ensure that the appropriate power cord feature is configured to support the power 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 save 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 models.

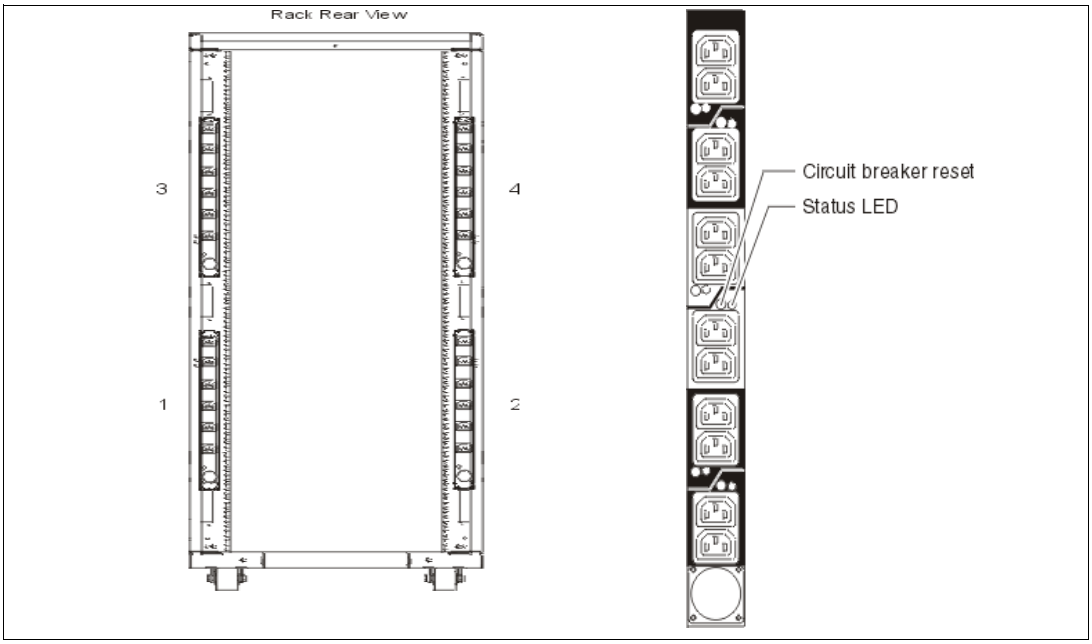


Figure 1-4 PDU placement and PDU view

1.5.7 Rack-mounting rules

The primary rules that should be followed when mounting the server into a rack are:

- The p5-520 or p5-520Q 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 a p5-520 or p5-520Q into the service position, it is essential that the rack manufacturer's safety instructions have been followed regarding rack stability.

The availability of 14 ft., 9 ft., and 6 ft. length jumper cords (between the drawer and the PDU) assist you in rack cable management by providing several options to ensure all cables are accounted for inside the rack space.

Depending on the current implementation and future enhancements of additional 7311 Model D20 drawers connected to the system, Table 1-10 shows examples of the minimum and maximum configurations for different combinations of servers and attached 7311 Model D20 I/O drawers.

Table 1-10 Minimum and maximum configurations for servers and 7311-D20s

	Only servers	One server, one 7311-D20	One server, four 7311-D20s
7014-T00 rack	9	4	1
7014-T42 rack	10	5	2
7014-S11 rack	2	1	0
7014-S25 rack	6	3	1

1.5.8 Additional options for rack

This section highlights some solutions available to provide a single point of management for environments composed of multiple System p5-520, p5-520Q servers or other IBM System p5 servers.

IBM 7212 Model 102 IBM TotalStorage storage device enclosure

The IBM 7212 Model 102 is designed to provide efficient and convenient storage expansion capabilities for selected System p servers. The IBM 7212 Model 102 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 102 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.
- Digital Data Storage (DDS-4) tape drive provides 20 GB native data capacity per tape cartridge and a native physical data transfer rate of up to 3 MBps that uses a 2:1 compression so that a single tape cartridge can store up to 40 GB of data.

- DVD-ROM drive is a 5 1/4-inch, half-high device. It can read 640 MB CD-ROM and 4.7 GB DVD-RAM media. It can be used for alternate IPL[3] (IBM-distributed CD-ROM media only) and program distribution.
- DVD-RAM drive with up to 2.7 MBps throughput. Using 3:1 compression, a single disk can store up to 28 GB of data. Supported DVD disk native capacities on a single DVD-RAM disk are as follows: up to 2.6 GB, 4.7 GB, 5.2 GB, and 9.4 GB.

Flat panel display options

The IBM 7316-TF3 Flat Panel Console Kit can be installed in the system rack. This 1U console uses a 15-inch thin film transistor (TFT) LCD with a viewable area of 304.1 mm x 228.1 mm and a 1024 x 768 pels[4] resolution. The 7316-TF3 Flat Panel Console Kit has the following attributes:

- Flat panel color monitor.
- Rack tray for keyboard, monitor, and optional VGA switch with mounting brackets.
- 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 Netbay LCM (Keyboard/Video/Mouse) Switch (FC 4202) provides users single-point access and control of up to 64 servers from a single console. The Netbay LCM Switch has a maximum video resolution of 1600x 280 and mounts in a 1U drawer behind the 7316-TF3 monitor. A minimum of one LCM feature (FC 4268) or USB feature (FC 4269) is required with a Netbay LCM Switch (FC 4202). 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.

[3] Initial program load
[4] Picture elements

Note: A server microcode update might be required on installed systems for boot-time System Management Services (SMS) menu support of the USB keyboards. The update might also be required for the LCM switch on the 7316-TF3 console (FC 4202). For microcode updates, see the following URL:

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

We recommend that you have the 7316-TF3 installed between EIA 20 to 25 of the rack for ease of use. The 7316-TF3 or any other graphics monitor requires a POWER GXT135P graphics accelerator (FC 2849) to be 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 Section 2.13, “Hardware Management Console” on page 54.

1.5.9 OEM rack

The p5-520 or p5-520Q 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 (17.75 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 1-5 on page 20 for a top view showing the specification dimensions.

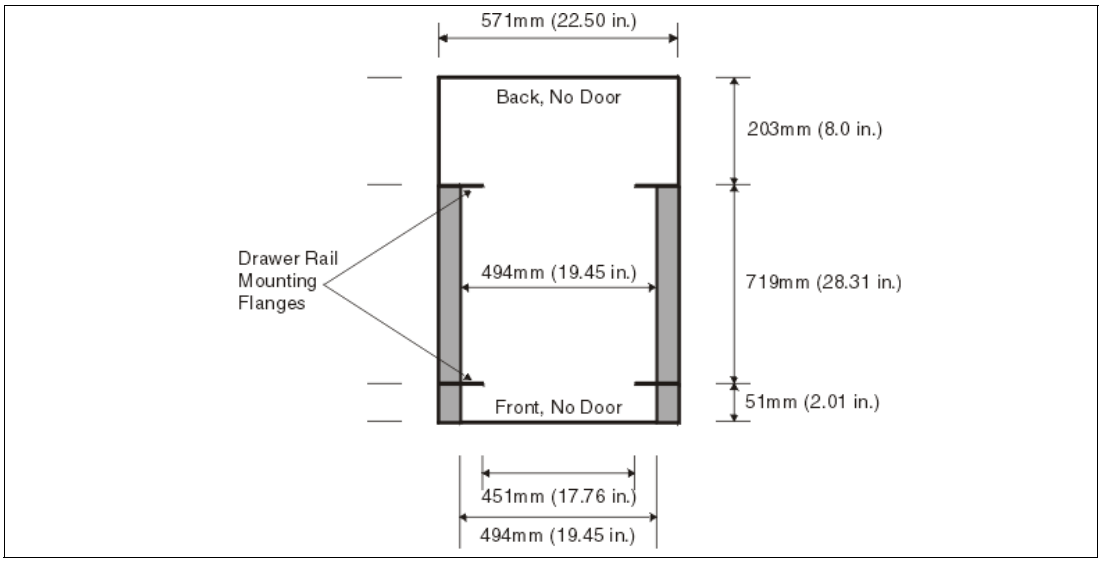


Figure 1-5 Top view of non-IBM rack specification dimensions

- 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 1-6 and Figure 1-7 on page 21 for the top and bottom front specification dimensions.

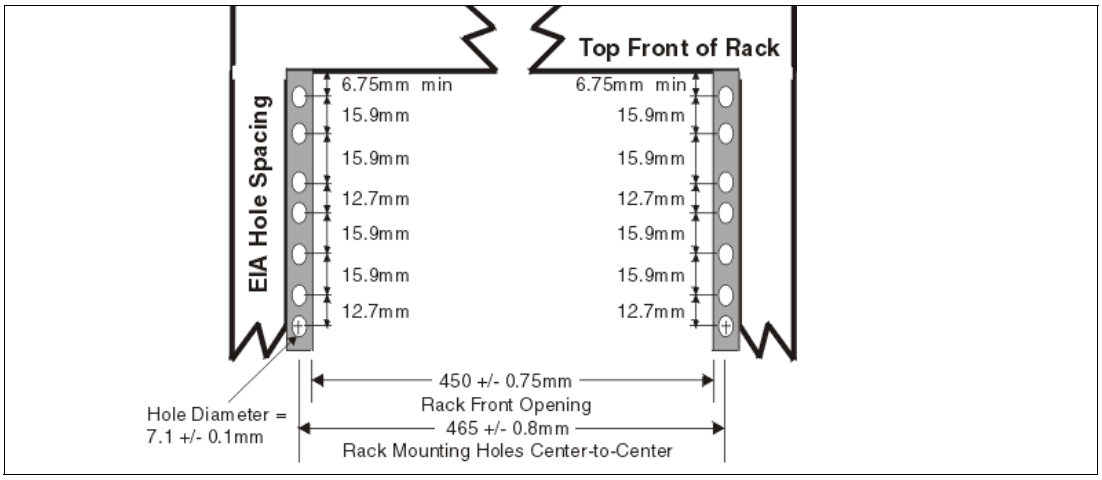


Figure 1-6 Rack specification dimensions, top front view

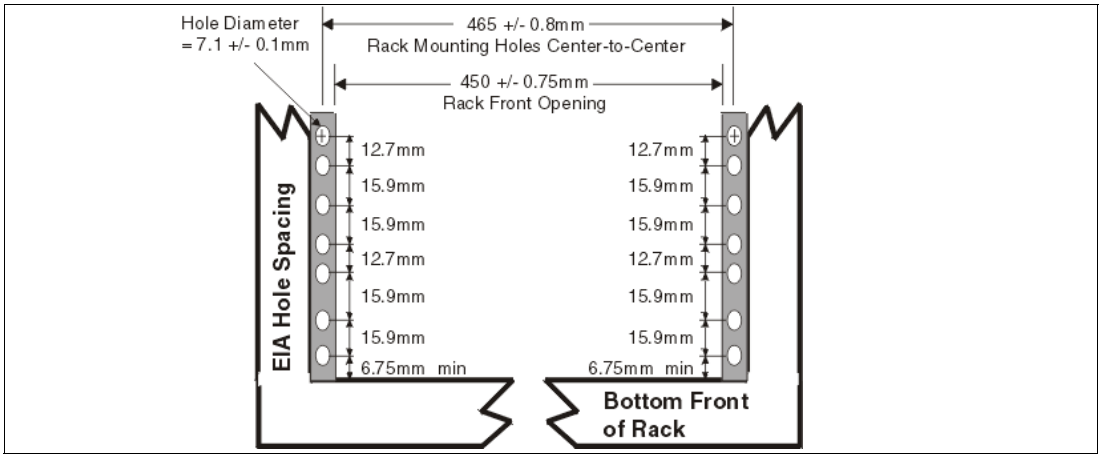


Figure 1-7 Rack specification dimensions, bottom front view

- 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 of the system p5-520 and p5-520Q. The architecture represented by the Figure 2-1 is detailing the base system hardware, and the DCM or QCM options. It is not allowed mixed install DCM and QCM. bandwidths provided throughout this section are theoretical maximums provided for reference. It is always recommended to obtain real-world performance measurements using production workloads.

\

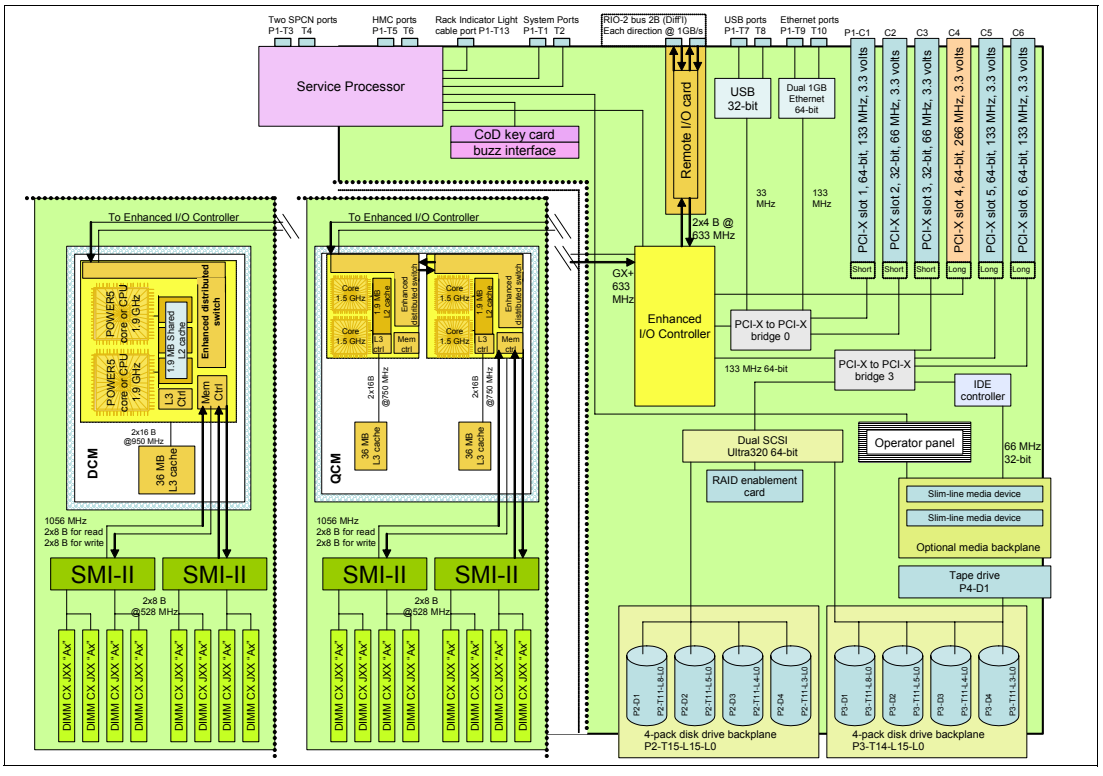


Figure 2.1 p5-520 and p5-520Q architecture with QCM or DCM

2.1 The POWER5+ chip

The IBM POWER5+ chip capitalizes all the enhancements brought by the POWER5 chip. For a detailed description of the POWER5 chip, refer to *IBM System p5-520 Technical Overview and Introduction,* REDP-9111.

Figure 2-1 shows a high level view of the POWER5+ chip.

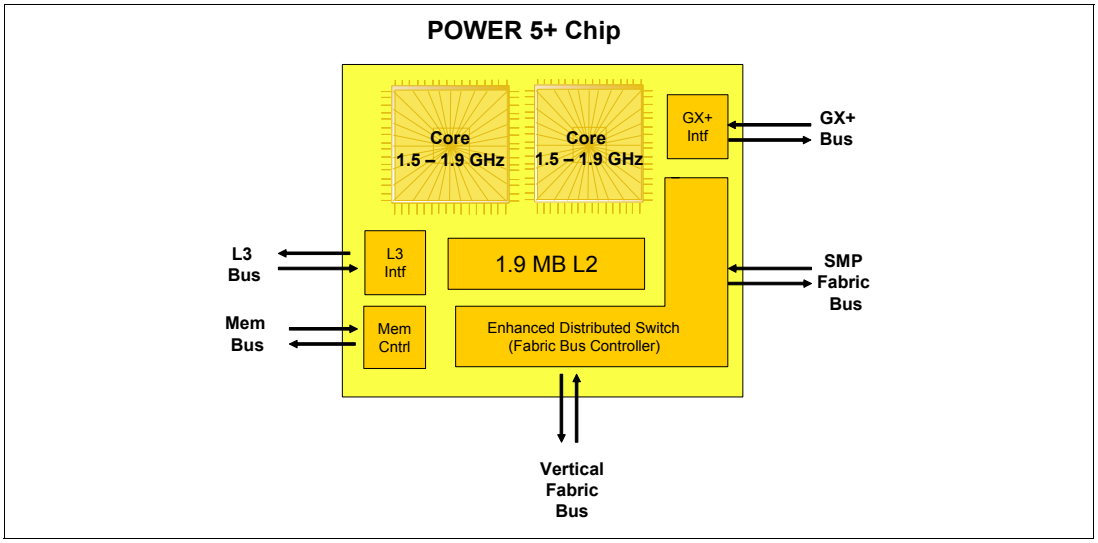


Figure 2-1 Power5+ chip

The CMOS9S technology used in the POWER5 chip used a 130 nm fabrication process. The CMOS10S technology for in the POWER5+ chip uses a 90 nm fabrication process, enabling:

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

Compared to the POWER5 chip, the 37 percent smaller POWER5+ chip consumes less power, thus requiring less sophisticated cooling. This allows its use in servers that previously only used low frequency chips due to cooling restrictions.

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

- New pages sizes in ERAT and TLB. Two new pages sizes (64 KB and 16 GB) recently added in PowerPC® architecture.
- New segment size in SLB. One new segment size (1 TB) recently added in PowerPC architecture.
- The TLB size has been doubled in the POWER5+ over the POWER5 processors. The TLB in POWER5+ has 2048 entries.
- Floating-point round to integer instructions. New instructions (frfin, frfiz, frfip, frfim) have been added to round floating-point numbers integers with the following rounding modes: nearest, zero, integer plus, integer minus.
- Improved floating-point performance.
- Lock performance enhancement.
- Enhanced SLB read.
- True Little-Endian mode. Support for the True Little-Endian mode as defined in the PowerPC architecture.

- Double the SMP support. Changes have been made in the fabric, L2 and L3 controller, memory controller, GX controller and chip RAS to provide support for the QCM that allows the SMP system sizes to be double that is available in POWER5 DCM-based servers. However current POWER5+ implementations only supports single address loop.
- Several enhancements have been made in the memory controller for improved performance. Ready to support for DDR2 667 MHz DIMMs in the future.
- Enhanced redundancy in L1 cache, L2 cache and L3 directory. Independent control of the L2 cache and the L3 directory for redundancy to allow split-repair action has been added. More word line redundancy has been added in the L1 Dcache. In addition, Array Built-In Self Test (ABIST) column repair for the L2 cache and the L3 directory has been added.

2.2 Processor and cache

In the p5-520 and p5-520Q, the POWER5+ chips, associated L3 cache chips, if any, and memory DIMMs are packaged on the system planar. The p5-520 1-core, 2-core and the p5-520Q 4-core use different POWER5+ processor modules.

Note: Because the 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 p5-520Q Quad-Core Module

The 4-core p5-520Q system planar contains a new Quad-Core Module (QCM) and the local memory storage subsystem for that QCM. Two POWER5+ dual core chips and their associated L3 cache chips are packaged in the QCM.

Figure 2-2 shows a layout view of a p5-520Q QCM with associated memory.

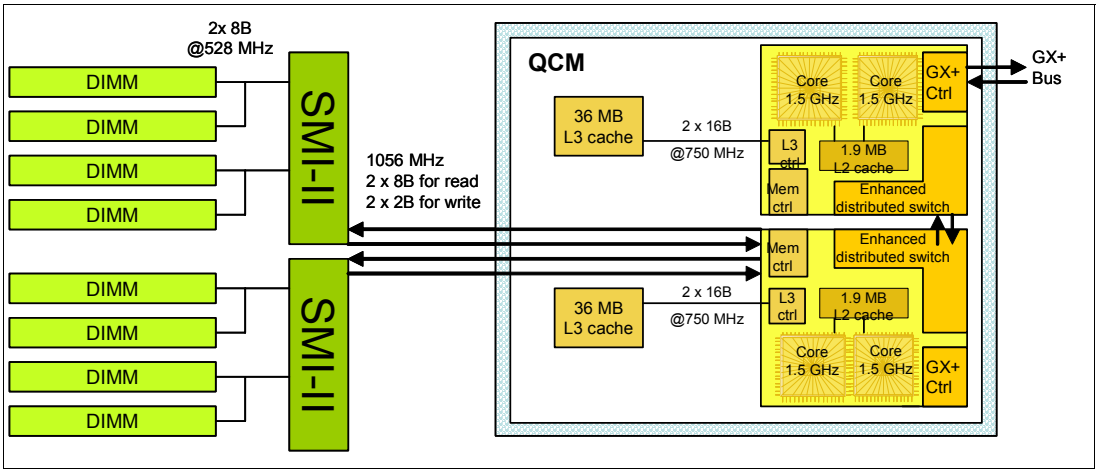


Figure 2-2 p5-520Q POWER5+ 1.5 GHz processor card with DDR2 memory socket layout view

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

I/O connects to the p5-520Q QCM using the GX+ bus. The QCM provides a single GX+ bus. Each processor in the POWER5+ chips 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+ chip that doesn't have direct access to memory does have a direct access to the GX+ Bus.

The theoretical maximum throughput of the L3 cache is 16 byte read, 16 byte write at a bus frequency of 750 MHz (based on a 1.5 GHz processor clock), which equates to 24000 MBps or 24 GBps.

2.2.2 p5-520 POWER5+ Dual-Core Module

The 2-core p5-520 system planar contains a Dual-Core Module (DCM) and the local memory storage subsystem for that DCM. The POWER5+ dual core chip and its associated L3 cache chip are packaged in the DCM.

Figure 2-3 shows a layout view of p5-520 DCM and associated memory.

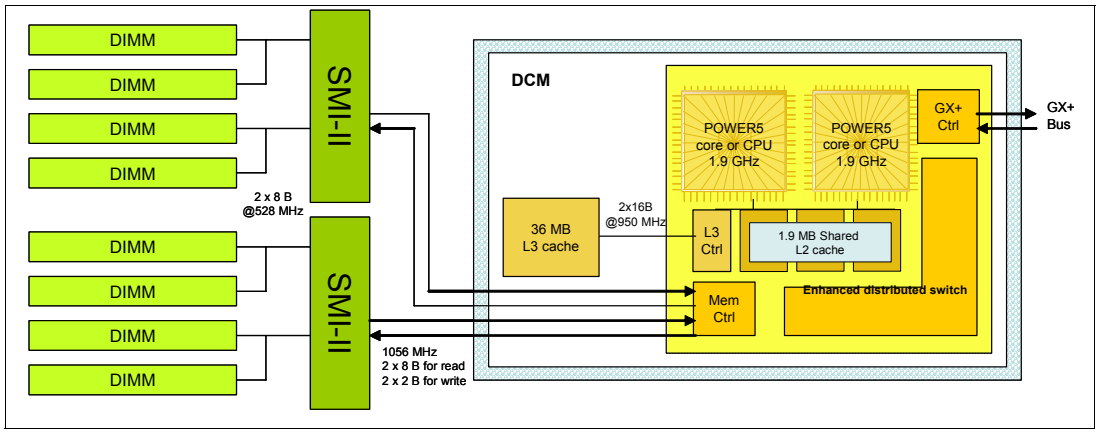


Figure 2-3 p5-520 POWER5+ 1.9 GHz DCM with DDR2 memory socket layout view

The storage structure for the POWER5+ chip 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, controlled by two SMI-II chips, which are located in close physical proximity to the processor modules.

I/O connects to the p5-520 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.

The theoretical maximum throughput of the L3 cache is 16 byte read, 16 byte write at a bus frequency of 950 MHz (based on a 1.9 GHz processor clock), which equates to 30400 MBps or 30.40 GBps.

2.2.3 p5-520 Single-Core Module

The 1-core p5-520 system planar contains a Single-Core Module (SCM) and the local memory storage subsystem for that SCM. L3 Cache is not available in this configuration.

2.2.4 Available processor speeds

The Table 2-1 describes what are the available processor capacities and speeds for the p5-520 and p5-520Q systems.

Table 2-1 p5-520 and p5-520Q available processor capacities and speeds

	p5-520 @ 1.9 GHz	p5-520 @ 1.65 GHz	p5-520Q @ 1.5 GHz
1-core	No	Yes	No
2-core	Yes	Yes	No
4-core	No	No	Yes

To determine the processor characteristics, 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 4-core p5-520Q system running at 1.5 GHz with simultaneous multithreading enabled:

```
Cpu 0 runs at 1498 MHz
Cpu 1 runs at 1498 MHz
Cpu 2 runs at 1498 MHz
Cpu 3 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 p5-520 and p5-520Q servers offer pluggable DDR2 DIMMs for memory. DDR2 DIMMs has double rate compared with DDR DIMMs (DDR DIMMs has double rate bits compared with SDRM) so that enables up to four times the performance of traditional SDRAM. The system planar provides eight slots for up to eight pluggable DDR2 DIMMs. The minimum memory for a p5-520 or p5-520Q server is 1.0 GB (2 x 512 MB) and 32 GB is the maximum installable memory option. Figure 2-4 on page 29 shows the memory slot and location codes. All memory is accessed by two of Synchronous Memory Interface (SMI)-II chips that are located between memory and processor. The SMI-II supports multiple data flow modes.

2.3.1 Memory placement rules

The memory features available at the time of writing for the p5-520 and p5-520Q servers are listed in Table 2-2.

Table 2-2 Available memory features

Feature code	Description
1930	1 GB (2 x 512 MB) DIMMs, 276-pin DDR2, 533 MHz SDRAM
1931	2 GB (2 x 1 GB) DIMMs, 276-pin DDR2, 533 MHz SDRAM
1932	4 GB (2 x 2 GB) DIMMs, 276-pin DDR2, 533 MHz SDRAM
1934	8 GB (2 x 4 GB) DIMMs, 276-pin DDR2, 533 MHz SDRAM

Memory can be pluggable in pairs or quads, as required by the total memory requirement. Memory feature numbers may be mixed within a system. The DIMMs slots are accessed by first removing the PCI riser book.

When additional memory is added to a system using FC 1930, an additional feature FC 1930 must be added to the original pair to make a quad, allowing one additional quad to be added to the system. Memory is installed in the first quad in the following order: J2A, J0A, J2C, and J0C; and for the second quad, in the order J2B, J0B, J2D, and J0D. Memory must be balanced across the DIMM quad slots. The Service Information label, located on the top cover of the system, provides memory DIMMs slot location information.

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 Fales
```

Note: A quad must consist of a single feature (that is, be made of identical DIMMs). Mixed DIMM capacities in a quad will result in reduced RAS.

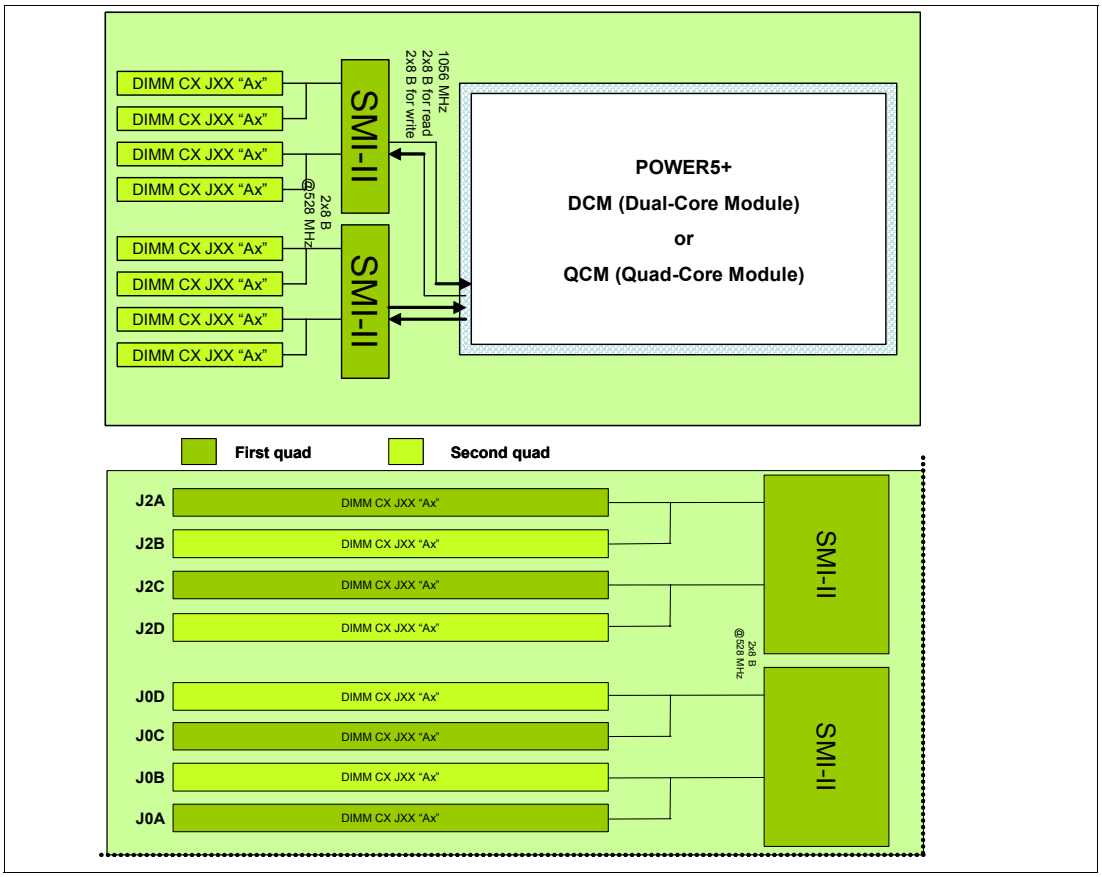


Figure 2-4 Memory placement for the p5-520 and p5-520Q servers

2.3.2 OEM memory

OEM memory is not supported or certified by IBM for use in a System p5 server. If the server is populated with OEM memory, you could experience unexpected and unpredictable behavior, especially when the system is using Micro-Partitioning technology.

All IBM memory is identified by an IBM logo and a white label printed with a barcode and an alphanumeric string, illustrated in Figure 2-5.

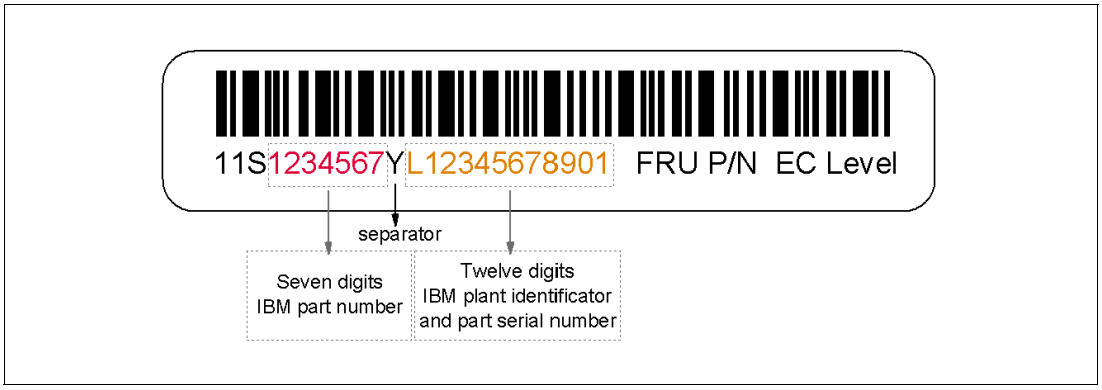


Figure 2-5 IBM memory certification label

2.3.3 Memory throughput

The memory subsystem throughput is based on the speed of the memory. An elastic interface, contained in the POWER5 chip, buffers reads and writes to and from memory and the processor. There are two Synchronous Memory Interface (SMI-II) chips, each with a single 8-byte read and 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 1066 MHz, the peak processor-to-memory throughput for read is (8 x 2 x 1056) = 16896 MBps or 16.89. The peak processor-to-memory throughput for write is (2 x 2 x 1056) = 4224 MBps or 4.22 GBps, making total 21.12 GBps.

The DIMMs are 533 MHz DDR2 operating 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.

2.4 I/O Buses

The following section provides additional information related to the internal buses.

2.4.1 RIO-2 buses and GX+ buses

The QCM or DCM provides the 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, connecting 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+ passthru port and four PCI-X 2.0 buses. The GX+ passthru 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).

The p5-520 and p5-520Q systems do not have RIO-2 ports integrated on the system planar to connect supported external I/O subsystems. As shown in Figure 2-6, one Remote I/O expansion card (FC 2888) is required to connect the supported external I/O subsystems. When this card is present, the Enhanced I/O Controller routes the GX+ bus to the external RIO-2 ports.

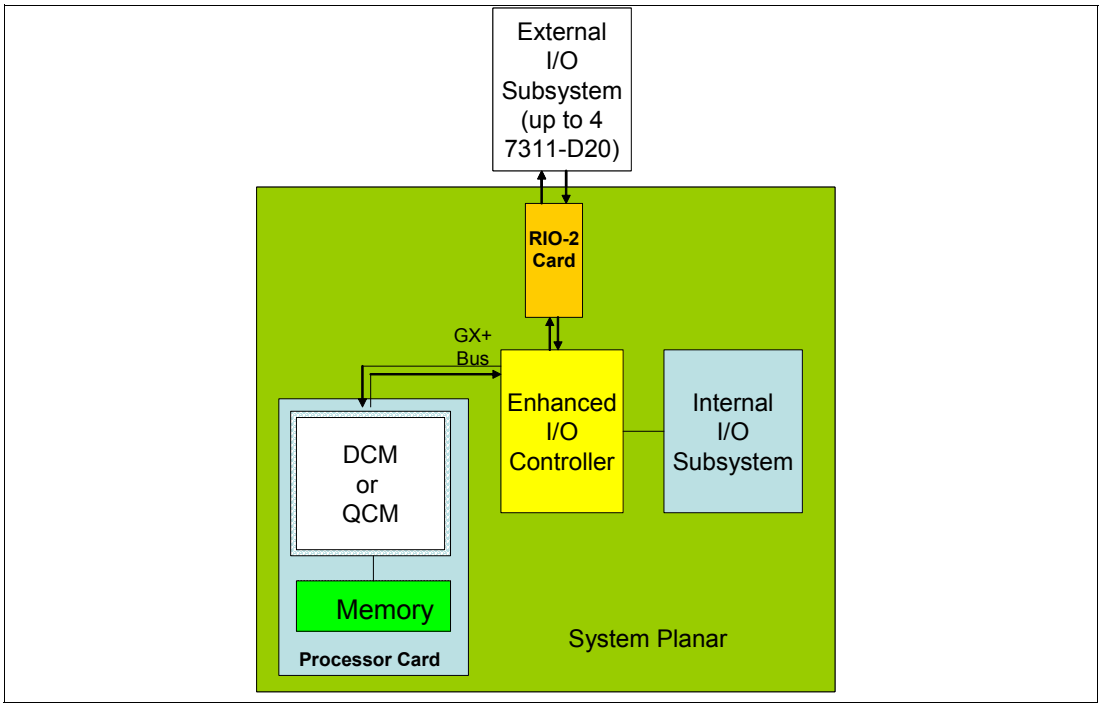


Figure 2-6 p5-520 or p5-520Q GX+ Bus connection overview

According to the processor speed, the I/O subsystem is capable to support 5.06 GBps when using the 1.9 GHz processor, or capable to support 4.0 GBps when using a 1.5 GHz processor. The bus is a dual four-byte wide bus running at a 3:1 processor to bus ratio.

2.5 Internal I/O subsystem

PCI-X, where the X stands for extended, is an enhanced PCI bus, delivering a bandwidth of up to 2 GBps, running a 64-bit bus at 133 MHz or 266 MHz. PCI-X is backward compatible, so the systems can support existing 3.3 volt PCI adapters.

The system planar provides six PCI-X slots and several integrated I/O devices. The PCI-X slot 1, slot 5 and slot 6 are 64-bit capable running at 133 MHz. The PCI-X slot 2 and slot 3 are 32-bit capable running at 66 MHz, but PCI-X 64-bit short adapters can be used in these slots.

All the PCI-X slots and the integrated I/O devices, except the PCI-X slot 4, are connected through two EADS-X chips that function as PCI-X to PCI-X bridges to the Enhanced I/O Controller. The connections of the PCI-X slots and integrated I/O devices to the PCI-X to PCI-X bridges are properly distributed to maximize the system performances.

The first three PCI-X slots can accept a short PCI-X or PCI card. The remaining PCI-X slots are full length cards. PCI-X slot 4 is a PCI-X DDR 266 MHz and 64 bit capable slot and is driven by the Enhanced I/O Controller directly. The dual 10/100/1000 Mbps Ethernet adapter and the Dual Channel SCSI Ultra320 adapter are some of the integrated devices on the system planar.

The PCI-X slots in the p5-520 and p5-520Q system support hot-plug and Extended Error Handling (EEH). In the unlikely event of a problem, EEH-enabled adapters respond to a special data packet generated from the affected PCI-X slot hardware by calling system firmware, which will examine the affected bus, allow the device driver to reset it, and continue without a system reboot.

2.6 64-bit and 32-bit adapters

IBM offers 64-bit adapter options for the p5-520 and p5-520Q, as well as 32-bit adapters. Higher-speed adapters use 64-bit slots because they can transfer 64 bits of data for each data transfer phase. Generally, 32-bit adapters can function in 64-bit PCI-X slots; however, some 64-bit adapters cannot be used in 32-bit slots. For a full list of the adapters that are supported on the systems, and for important information regarding adapter placement, see the *IBM Systems Hardware Information Center*. You can find it at:

http://publib.boulder.ibm.com/eserver/

The internal PCI-X slots support a wide range of PCI-X I/O adapters to handle your I/O requirements.

2.6.1 LAN adapters

To connect a p5-520 or p5-520Q to a local area network (LAN), the dual port internal 10/100/1000 Mbps RJ-45 Ethernet controller, integrated on the system planar can be used.

See Table 2-3 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 (CHRP[1] is the platform type). token-ring is not allowed as initial order.

Table 2-3 Available LAN adapters

Feature code	Adapter description	Type	Slot	Size	Max
1954	4-port 10/100/1000 Ethernet	Copper	32 or 64	short	4
1978	Gigabit Ethernet	Fibre	32 or 64	short	5
1979	Gigabit Ethernet	Copper	32 or 64	short	5
1981	10 Gigabit Ethernet - short reach	Fibre	32 or 64	short	2
1982	10 Gigabit Ethernet - long reach	Fibre	32 or 64	short	2
1983	2-port Gigabit Ethernet	Copper	32 or 64	short	5
1984	2-port Gigabit Ethernet	Fibre	32 or 64	short	5

2.6.2 SCSI adapters

To connect to external SCSI devices, the adapters provided in Table 2-4 are available, at the time of writing, to be configured with an initial order.

Table 2-4 Available SCSI adapters

Feature code	Adapter description	Slot	Size	Max
1912	Dual Channel Ultra320 SCSI	64	short	5
1913	Dual Channel Ultra320 SCSI RAID	64	long	2

Note: Previous SCSI adapters are also supported to be used in the p5-520 and p5-520Q but cannot be part of an initial order configuration. Clients that would like to connect existing external SCSI devices can contact the IBM service representative.

[1] CHRP stands for Common Hardware Reference Platform, a specification for PowerPC-based systems that can run multiple operating systems.

There is also the option to make the internal Ultra320 SCSI channel externally accessible on the rear side of the system by installing FC 4275. No additional SCSI adapter is required in this case. If FC 4275 is installed, a second 4-pack disk enclosure (FC 6574 or FC 6594) cannot be installed, which limits the maximum number of internal disks to four. FC 4275 also requires one PCI-X slot.

For more information about the internal SCSI system, see 2.7, “Internal storage” on page 37.

2.6.3 Integrated RAID options

The p5-520 and p5-520Q can be configured with the optional SCSI RAID daughter card (FC 1907) that plugs directly on the system board or with a Dual Channel Ultra320 SCSI RAID adapter (FC 1913) to drive one 4-pack disk enclosure.

RAID implementation requires a minimum of three disk drives to form a RAID set.

These are different internal RAID options that may be considered:

- Install FC 1907 and up to 4 disk drives in the default 4-pack disk enclosure. This will allow RAID capabilities within a single 4-pack.
- Install FC 1907 and a second 4-pack disk enclosure (FC 6574). This will allow RAID capabilities across two 4-packs.
- Install FC 1907 and the optional 4-pack disk enclosure for disk mirroring (FC 6594). Install FC 1912, the PCI-X Dual Channel Ultra320 SCSI RAID adapter and the SCSI cable (FC 4267) which connects the PCI-X adapter to the optional 4-pack disk enclosure. This RAID configuration will provide increased reliability over first and second options
- Install the Dual Channel SCSI RAID Enablement Card (FC 1913). Install four disk drives in the first 4-pack DASD backplane (FC 6574). This will allow RAID 0, 5, or 10 capabilities within a single 4-pack of DASD with one RAID controller.
- Install FC 1913. Install a second FC 6574. Install four additional disk drives in the second 4-pack DASD backplane. This will allow RAID 0, 5, or 10 capabilities across two 4-packs of DASD with one RAID controller.
- Install FC 1913. Install the Ultra320 SCSI 4-Pack Enclosure for Disk Mirroring (FC 6594). Install the PCI-X Dual Channel Ultra320 SCSI RAID Adapter (FC 1975). Install the SCSI Cable, which connects the PCI Adapter to the second 4-pack DASD backplane (FC 4267). This will allow RAID 0, 5, or 10 capabilities within each 4-pack of DASD with two RAID controllers.

Note: Because the p5-520 and p5-520Q have up to eight disk drive slots, clients performing upgrades must perform appropriate planning to ensure the correct handling of their RAID arrays.

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. iSCSI allows an existing network to transfer SCSI commands and data with full location independence and defines 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, iSCSI supports Gigabit Ethernet and its frames so that systems supporting 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 to the operating system 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 in System p5 systems support iSCSI in the following two modes:

Hardware Using iSCSI adapters (see “IBM iSCSI adapters” on page 34)

Software Supported on standard Gigabit adapters, additional software (see “IBM iSCSI software Host Support Kit” on page 35) must be installed. The main processor is utilized for processing related to iSCSI protocol

Initial iSCSI implementations are targeted at 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, as it is typically in a data center, it can be accessed by the iSCSI SANs (and vice versa) via 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 reside 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 provide advantage of the increased bandwidth through the hardware support of iSCSI protocol. The 1 Gigabit iSCSI TOE PCI-X adapters support hardware encapsulation of SCSI commands and data into TCP and transports over the Ethernet using IP packets. The adapter operates as an iSCSI TOE (TCP/IP Offload Engine). This offload function eliminates host protocol processing and reduces CPU interrupts. Adapter uses Small form factor LC type fiber optic connector or copper RJ45 connector.

Table 2-5 provides the orderable iSCSI adapters.

Table 2-5 Available iSCSI adapters

Feature code	Description	Slot	Size	Max
1986	Gigabit iSCSI TOE PCI-X on copper media adapter	64	short	3
1987	Gigabit iSCSI TOE PCI-X on optical media adapter	64	short	3

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 software iSCSI initiator and allows to access iSCSI target storage devices using standard Gigabit Ethernet network adapters. To ensure the best performance enable the TCP Large Send, TCP send and receive flow control, and Jumbo Frame features of the Gigabit Ethernet Adapter and the iSCSI Target. Tune network options and interface parameters for maximum iSCSI I/O throughput on the operating system.

IBM System Storage N series

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

All models provide:

- Support for entry-level and midrange customers requiring 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 p5-520 and p5-520Q support the 2 Gigabit Fibre Channel PCI-X Adapter (FC 1977). The PCI-X adapter is a 64-bit, short form factor adapter with an LC type external fibre connector that provides single or dual initiator capability over an optical fiber link or loop. With the use of appropriate optical fiber cabling, this adapter provides the capability for a network of high speed local and remote located storage. Distances of up to 500 meters running at 1 Gbps data rate and up to 300 meters running at 2 Gbps data rate are supported between the adapter and an attaching device or switch. When used with IBM supported Fibre Channel storage switches supporting long-wave optics, distances of up to 10 kilometers are capable running at either 1 Gbps or 2 Gbps data rates.

The 2 Gigabit Fibre Channel PCI-X Adapter can be used to attach devices either directly, or using the supported Fibre Channel Switches. If attaching a device or switch with a SC type fibre connector, also the LC-SC 50 Micron Fiber Converter Cable (FC 2456) or a LC-SC 62.5 Micron Fiber Converter Cable (FC 2459) is required.

2.6.6 Graphic accelerators

The p5-520 and p5-520Q support up to four 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, as well as when AIX 5L or Linux starts an X11-based graphic user interface (GUI).

2.6.7 InfiniBand Host Channel adapter

The p5-520 and p5-520Q support the RIO-2 expansion cards (FC 2888) to connect the supported additional I/O subsystems. The server also supports one GX Dual-port 4x InfiniBand Host Channel Adapter (FC 1812) that enables the attachment of the Topspin Server Switch models 120 and 270. The GX Dual-port 4x InfiniBand HCA, as well as the RIO-2 expansion card if present, plugs into the system planar, using the GX slot. Connection to the Topspin Server Switches are accomplished by using the 4x IB Cables.

The systems also support up to two PCI-X Dual-port 4x InfiniBand Host Channel Adapter (FC 1820) to enable the attachment of the Topspin Server Switch as FC 1812.

Topspin Server Switch models 120 and 270

Switches are the fundamental components of an InfiniBand fabric. An IBM System p5 server proposal may also include the Topspin Server Switch model 120 and 270 in an initial system order.

The Topspin 120 and 270 Server Switch are a programmable switching platform that consists of a switched multiple-terabit interconnect and an intelligent control architecture. The high-bandwidth, low-latency interconnection is extremely adaptable. The switches enable an outstanding level of application scaling, rapid deployment, and resource consolidation.

See the following link for more Topspin Server Switch information:

http://www.topspin.com/solutions/index.htm

2.6.8 Asynchronous PCI-X adapters

The asynchronous PCI-X adapters provide connection of asynchronous EIA-232 or RS-422 devices. In case of a cluster configuration or high-availability configuration, if the plan is to connect the IBM System p5 servers using a serial connection, it is not supported to use the two default system ports but one of the following features is required:

Table 2-6 Asynchronous PCI-X adapters

Feature code	Description
2943	8-Port Asynchronous Adapter EIA-232/RS-422
5723	2-Port Asynchronous IEA-232 PCI Adapter

2.6.9 Additional support for owned PCI-X adapters

The lists of the major PCI-X adapters that can be configured in a p5-520 or p5-520Q when an initial configuration order is going to be built are described in the previous sections, from 2.6.1, “LAN adapters” on page 32 to 2.6.8, “Asynchronous PCI-X adapters” on page 36, but 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 owned PCI-X adapters can contact the IBM service representative to verify if supported.

2.6.10 Internal system ports

The system ports S1 and S2, at the rear of the system, are only available if the system is not managed using a Hardware Management Console (HMC). In this case, the S1 and S2 ports support the attachment of serial console and modem.

If an HMC is connected, a *virtual serial console* is provided by the HMC (logical device vsa0 under AIX 5L) and also 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 2.6.8, “Asynchronous PCI-X adapters” on page 36.

2.7 Internal storage

There is one dual channel Ultra320 SCSI controller managed by the EADS-X chips, integrated into the system planar, that are used to drive the internal disk drives. The eight internal drives plug into the disk drive backplane, which has two separate SCSI buses with four disk drives per bus.

The internal disk drive can be used in two different modes based on whether the SCSI RAID Enablement Card (FC 1976) is installed (see 2.6.3, “Integrated RAID options” on page 33).

The p5-520 and p5-520Q supports two 4-pack disk drives using a backplane that is designed for hot-pluggable disk drives. The disk drive backplane docks directly to the system planar. The virtual SCSI Enclosure Services (VSES) hot-plug control functions are provided by the Ultra320 SCSI controllers.

2.7.1 Internal media devices

The p5-520 and p5-520Q provides two slim-line media bays for optional DVD-ROM (FC 1994) and optional DVD-RAM (FC 1993), and one media bay for a tape drive. Table 2-7 shows all additional media devices for the systems.

Table 2-7 Available tape drives

Feature code	Description
1892	VXA-320 160/320 GB Internal Tape Drive
1991	36/72 GB 4 mm Internal Tape Drive
1992	IBM 80/160 GB Internal Tape Drive with VXA Technology
1997	200/400 GB Half High Ultrium 2 Tape Drive

2.7.2 Internal hot-swappable SCSI disks

The p5-520 and p5-520Q can have up to eight hot-swappable disk drives plugged in the two 4-pack disk drives backplanes. The hot-swap process is controlled by the SCSI enclosure service (SES), which is located in the 4-pack disk drives backplane (AIX 5L assigns the name `ses0` to the first 4-pack, and `ses1` to the second, if present). The two hot-swappable 4-pack disk drives backplanes can accommodate the devices listed in Table 2-8.

Table 2-8 Available hot-swappable disk drives

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

Feature code	Description
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

At the time of writing, if a new order is placed with two 4-pack DASD backplanes (FC 6574) and more than one disk, the system configuration shipped from manufacturing will balance the total number of SCSI disks between the two 4-pack SCSI backplanes. This is for manufacturing test purposes, and not because of any limitation. Having the disks balanced between the two 4-pack DASD backplanes allows the manufacturing process to systematically test the SCSI paths and devices related to them.

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

Note: We recommend that you follow this procedure, after the disk has been deconfigured, when removing a hot-swappable disk:

1. Release the tray handle on the disk.
2. Pull out the disk assembly a little bit from the original position.
3. Wait up to 20 seconds until the internal disk stops spinning.
4. Now you can safely remove the disk from the 4-pack DASD backplane.

After the SCSI disk hot-swap procedure, you can expect to find `SCSI_ERR10` logged in the AIX 5L error log, with the second word of the sense data equal to `0017`. It is generated from a SCSI bus reset issued by the SES to reset all processes when a drive is inserted, and it is not an issue.

Hot-swappable disks and Linux

Hot-swappable disk drives on IBM System p5 systems are supported with SUSE Linux Enterprise Server 9 for POWER, or later, and Red Hat Enterprise Linux AS for POWER Version 3, or later.

2.8 External I/O subsystem

This section describes the external I/O subsystem, the 7311 D20 I/O drawer that is the only drawer supported on the p5-520 and p5-520Q systems.

2.8.1 I/O drawers

As described in Chapter 1, “General description” on page 1, the p5-520 or p5-520Q systems have six internal PCI-X slots, which is enough in many cases. If more PCI-X slots are needed to dedicate more adapters to a partition or to increase the bandwidth of network adapters, up to four 7311 Model D20 I/O drawers can be added to the p5-520 or p5-520Q systems.

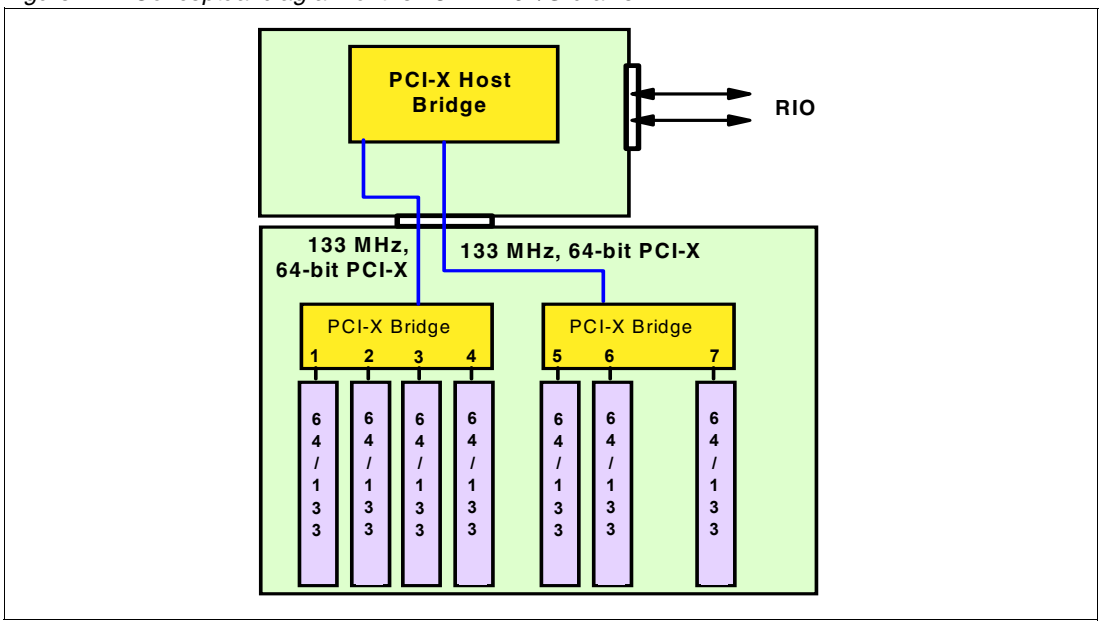
The p5-520 or p5-520Q systems have a standard RIO-2 bus to connect the internal PCI-X slots through the PCI-X to PCI-X bridges and support up to four external I/O drawers.

The 7311 Model D20 I/O drawer must have the RIO-2 loop adapter (FC 6417) to be connected to the p5-520 or p5-520Q systems. The PCI-X host bridge inside the I/O drawer provides two primary 64-bit PCI-X buses running at 133 MHz. Therefore, a maximum bandwidth of 1 GBps is provided by each of the buses. To avoid overloading an I/O drawer, the recommendation in the IBM System p5 Hardware Information Center should be followed. You can find it at:

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

Figure 2-7 shows a conceptual diagram of the 7311 Model D20 I/O drawer subsystem.

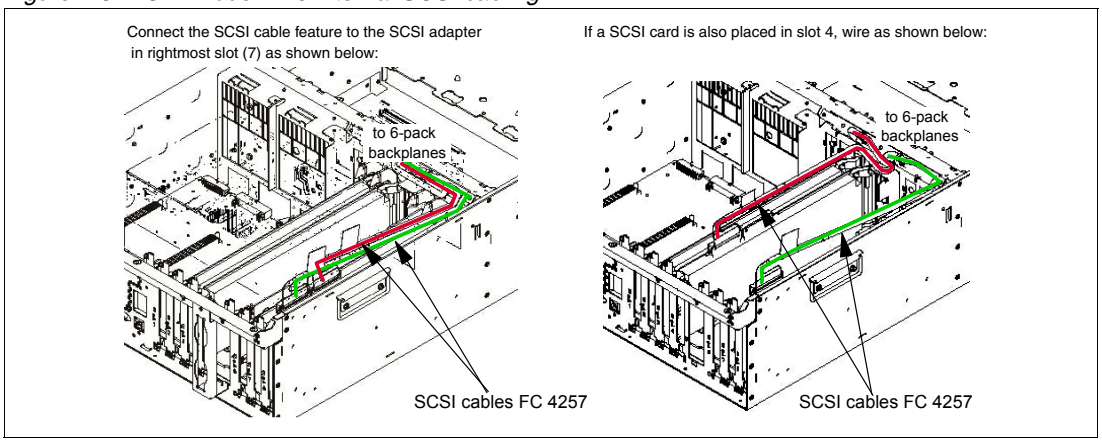
Figure 2-7 Conceptual diagram of the 7311-D20 I/O drawer



7311 Model D20 internal SCSI cabling

A 7311 Model D20 supports hot-swappable disks using two 6-pack disk bays for a total of 12 disks. Additionally, the SCSI cables (FC 4257) are used to connect a SCSI adapter (that can have various features) in slot 7 to each of the 6-packs, or two SCSI adapters, one in slot 4 and one in slot 7 (see Figure 2-8).

Figure 2-8 7311 Model D20 internal SCSI cabling



Note: Any 6-packs and the related SCSI adapter can be assigned to a partition. If one SCSI adapter is connected to both 6-packs, both 6-packs can be assigned only to the same partition. When the server is configured with the The Advanced POWER Virtualization hardware feature and the Virtual I/O Server is used for virtual SCSI, the disks can be shared between partitions.

2.8.2 7311 I/O drawer RIO-2 cabling

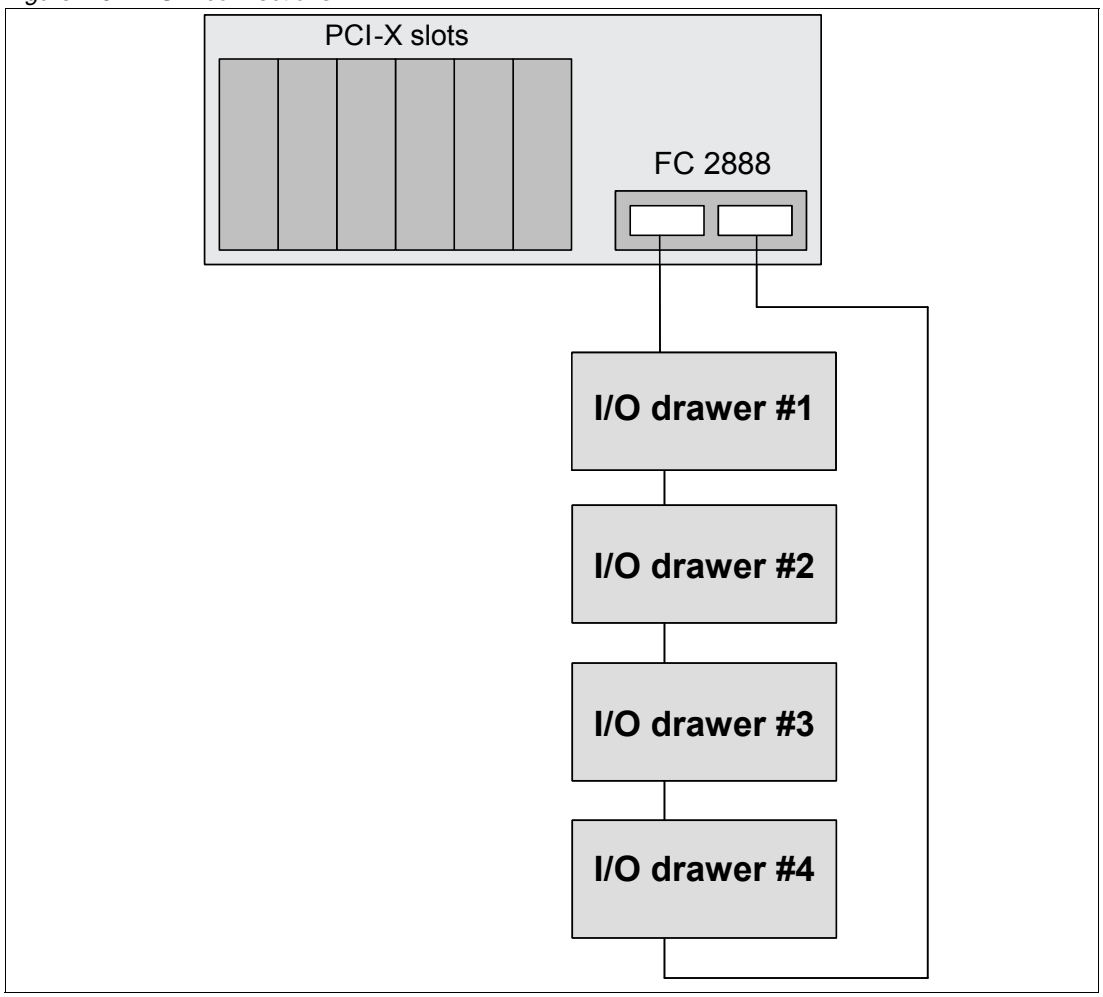
As described in 2.8, “External I/O subsystem” on page 38, you can connect up to four I/O drawers in the same loop to the p5-520 or p5-520Q system.

Each RIO-2 port can operate at 1 GHz in bidirectional mode and is capable of passing data in each direction on each cycle of the port. Therefore, the maximum data rate is 4 GBps per I/O drawer in double barrel mode.

There is one default primary RIO-2 loop in any p5-520 or p5-520Q system. This feature provides two Remote I/O ports for attaching up to four 7311 Model D20 I/O drawers to the system in a single loop.

Figure 2-9 shows how you could connect four I/O drawers to one system.

Figure 2-9 RIO-2 connections



The RIO-2 cables used have different lengths to satisfy the different connection requirements:

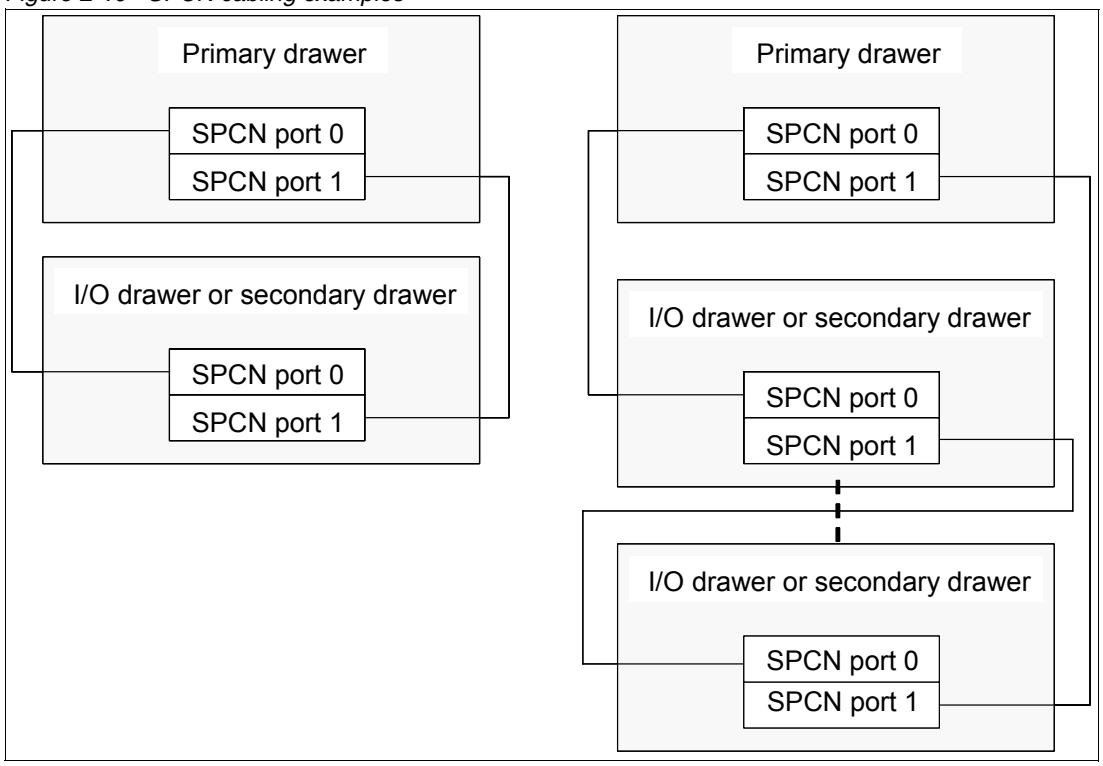
- Remote I/O cable, 1.2 m (FC 3146)
- Remote I/O cable, 1.75 m (FC 3156)
- Remote I/O cable, 2.5 m (FC 3168)
- Remote I/O cable, 3.5 m (FC 3147)
- Remote I/O cable, 10 m (FC 3148)

2.8.3 7311 Model D20 I/O drawer SPCN cabling

The SPCN is used to control and monitor the status of power and cooling within the I/O drawer. The SPCN is a loop, the cabling starts from SPCN port 0 on the p5-520 or p5-520Q system to SPCN port 0 on the first I/O drawer. The loop is closed connecting the SPCN port 1 of the I/O drawer back to the port 1 of p5-520 or p5-520Q system. If you have more than one I/O drawer, you continue the loop connecting the following drawer (or drawers) with the same rule.

See Figure 2-10 for SPCN cabling examples.

Figure 2-10 SPCN cabling examples



There are different SPCN cables to satisfy different length requirements:

- SPCN cable drawer to drawer, 2 m (FC 6001)
- SPCN cable drawer to drawer, 3 m (FC 6006)
- SPCN cable rack to rack, 6 m (FC 6008)
- SPCN cable rack to rack, 15 m (FC 6007)
- SPCN cable rack to rack 30 m (FC 6029)

2.9 External disk subsystems

The p5-520 and p5-520Q have internal hot-swappable drives. When the AIX 5L operating system is installed in a IBM System p5 servers, the internal disks are usually used for the AIX 5L rootvg volume group and paging space. Specific client requirements can be satisfied with the several external disk possibilities that the system supports.

2.9.1 IBM TotalStorage EXP24 Expandable Storage

The IBM TotalStorage® EXP24 Expandable Storage disk enclosure, Model D24 or T24, can be purchased together with the p5-520 or p5-520Q and will provide 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 deskside (Model T24) unit. Whether high availability storage solutions or simply high capacity storage for a single server installation, the unit provides a cost-effective solution. It provides 24 hot-swappable disk bays, 12 accessible from the front and 12 from the rear. Disk options that can be accommodate in any of the four six-packs disk drive enclosure are 73.4 GB, 146.8 GB or 300 GB 10K rpm or 36.4 GB, 73.4 GB or 146.8 GB 15K rpm drives. Each of the four six-packs disk drive enclosure may 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, installed in one or multiple servers or logical partitions. Optionally, the two front or two rear six-packs may be connected together to form a single Ultra320 SCSI bus of 12 drives.

2.9.2 IBM System Storage N3000 and N5000

The IBM System Storage N3000 and N5000 line of iSCSI enabled storage offerings provide a flexible way to implement a Storage Area Network over an Ethernet network.

2.9.3 IBM TotalStorage DS4000 Series

The IBM System Storage DS4000 line of Fibre Channel enabled Storage offerings provides a wide range of storage solutions for your Storage Area Network. The IBM TotalStorage DS4000 Storage server family consists of the following models: DS4100, DS4300, DS4500, and DS4800. The Model DS4100 Express 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 storage area network (SAN). If only space for the rootvg is needed, the Model DS4100 is a good solution.

For 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`

2.9.4 IBM TotalStorage Enterprise Storage Server

The IBM TotalStorage Enterprise Storage Server® (ESS) Models DS6000 and DS8000 are the high-end premier storage solution for use in storage area networks and use POWER technology-based design to provide fast and efficient serving of data. 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; however, the 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

2.10 Logical partitioning

Dynamic logical partitions (LPARs) and virtualization increase utilization of system resources and adds 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 on the system, such as the POWER Hypervisor, and optional ones, such as the Advanced POWER Virtualization feature.

2.10.1 Dynamic logical partitioning

Logical partitioning (LPAR) 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, allowing each LPAR to 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 executing. 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 dynamically redefine all available system resources to reach the optimum capacity for each defined dynamic LPAR.

Operating system support for dynamic LPAR

Table 2-9 lists AIX 5L and Linux support for dynamic LPAR capabilities.

Table 2-9 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

2.11 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 on System p5 and ℮server p5 servers.

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

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

and the following IBM Redbooks:

http://www.redbooks.ibm.com/abstracts/sg247940.html?Open
http://www.redbooks.ibm.com/abstracts/sg245768.html?Open

2.11.1 POWER Hypervisor

Combined with features designed into the POWER5 and POWER5+ processors, the POWER Hypervisor delivers functions that enable other system technologies, including Micro-Partitioning technology, virtualized processors, IEEE 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 provides the following functions:

- 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 later discussion in 2.12.2, “Logical, virtual, and physical processor mapping” on page 47).
- Saves and restores all processor state information during logical processor context switch.
- Controls hardware I/O interrupt management facilities for logical partitions.
- Provides virtual LAN channels between physical partitions that help to reduce the need for physical Ethernet adapters for inter-partition communication.

The POWER Hypervisor is always active when the server is running partitioned or not and also when not connected to the HMC. It requires memory to support the logical partitions on the server. The amount of memory required by the POWER Hypervisor firmware varies according to several factors. Factors influencing the POWER Hypervisor memory requirements include the following:

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

Note: Use the LPAR Validation Tool for estimate the memory requirements of the POWER Hypervisor.

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

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 may 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 virtual SCSI mechanism for virtualization of storage devices (a special logical partition to install the Virtual I/O Server is required to utilize this feature, see 2.12.3, “Virtual I/O Server” on page 49). The storage virtualization is accomplished using 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 7940).

Virtual Ethernet

The POWER Hypervisor provides a virtual Ethernet switch function that allows partitions on the same server to use a fast and secure communication without any need for physical interconnection. The virtual Ethernet allows a transmission speed in the range of 1 to 3 GBps. depending on the MTU[2] size and CPU entitlement. Virtual Ethernet requires system with either AIX 5L Version 5.3 or appropriate level of Linux supporting virtual Ethernet devices (see chapter 2.14, “Operating system support” on page 57). 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 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 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 imply that each Virtual Ethernet adapter can be used to access 21 virtual networks.
- Each partition operating system detects the virtual local area network (VLAN) switch as an Ethernet adapter without the physical link properties and asynchronous data transmit operations.

Any virtual Ethernet can also have connection outside of the box if a layer-2 bridging to a physical Ethernet adapter is set in one Virtual I/O Server partition (see “Virtual I/O Server” on page 49 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.

[2] Maximum transmission unit

Depending on the system configuration, the operating system console can be provided by the Hardware Management Console virtual TTY, IVM virtual TTY or from a terminal emulator connected to a system port.

2.12 Advanced POWER Virtualization feature

The Advanced POWER Virtualization feature (FC 7940) 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 will be enabled for micro-partitions (number of processors on 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 Version 5.3 or Linux. It supports:
  - Ethernet adapter sharing (Ethernet bridge from virtual Ethernet to external network).
  - Virtual SCSI Server.
  - Partition management using Integrated Virtualization Manager (Virtual I/O Server Version 1.2 or later only).
- Partition Load Manager (AIX 5L Version 5.3 only)
  - Automated CPU and memory reconfiguration.
  - Real-time partition configuration and load statistics.
  - Graphical user interface.

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

2.12.1 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 Section 2.12.2, “Logical, virtual, and physical processor mapping” on page 47 for more details.

When defining 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, the partition will be dispatched on physical processors.
- The processing sharing mode, either capped or uncapped.
- Weight (preference) in the case of uncapped partition.
- Minimum, desired, and maximum number of virtual processors.

POWER Hypervisor calculates a partition's processing *entitlement* based on minimum, desired and maximum values, sharing mode and also based on other active partitions' requirements. The actual etitlement is never smaller than the minimum value but can exceed the maximum value is case of 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 shared processor partitions and partition's entitlement 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 under 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 writing this publication and all shared partitions are dispatched by Hypervisor within this pool. Dedicated partitions and Micro-partitions can coexist on the same POWER5+ processor-based server as long as enough processors are available.

The systems support up to 4-core processor configuration, therefore up to four dedicated partitions, or up to 40 micro-partitions can be created. It is important to point out that the maximums stated are supported by the hardware, but the practical limits depend on the application workload demands.

2.12.2 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 a 4-core configuration with a QCM (Quad-Core Module) there are four physical processors.

In dedicated mode, physical processors are assigned as a whole to partitions. 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. Operating system (AIX 5L or Linux) sees one physical processor as two logical processors if the simultaneous multithreading feature is on. It can be turned off while 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 the time.

In a micro-partitioned environment with shared mode partitions an additional concept of *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 partition's processing units entitlement). At the end of the POWER Hypervisor's dispatch cycle, 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). Operating system is able to dispatch its software threads to these virtual processors and is completely screened from 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 virtual processor, the concept of logical processor exists and the number of logical processor depends whether the simultaneous multithreading is turned on or off.

Some additional information related to the virtual processors:

- There is one-to-one mapping of running virtual processors to physical processors at any given time. No more virtual processors can be active at any given time than the total number of physical processors in shared processor pool is.
- A virtual processor can be either running (dispatched) on a physical processor or standby waiting for a physical processor to become available.
- Virtual processors do not introduce any additional abstraction level, they are really only dispatch entity. When running on a physical processor they run at full speed of physical processor.
- Each partition's profile defines CPU entitlement that determines how much processing power any given partition should receive. Total sum of CPU entitlement of all partitions cannot exceed number of available physical processors in shared processor pool.
- A partition will have amount of processing power regardless of number of virtual processors it defines.
- A partition can use more processing power, regardless its entitlement, if it is defined as an *uncapped* partition in the partition profile. If there is spare processing power available in 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 in uncapped, the number of defined virtual processors determines the limitation of the maximum processing power it can receive. For example if number of virtual processors is two, then the maximum usable processor units is two.
- It is allowed to define more virtual processors than physical processors. In that case, 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 may benefit in using more virtual processors than physical processors.
- The number of virtual processors can be dynamically changed 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 the partition could ever request. If there are, for example, four physical processors installed in the system, one production partition and three test partitions, then:

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

Figure 2-11 on page 49 shows logical, virtual, and physical processor mapping, and an example of how the virtual processor and logical processor may be dispatched to the physical processor.

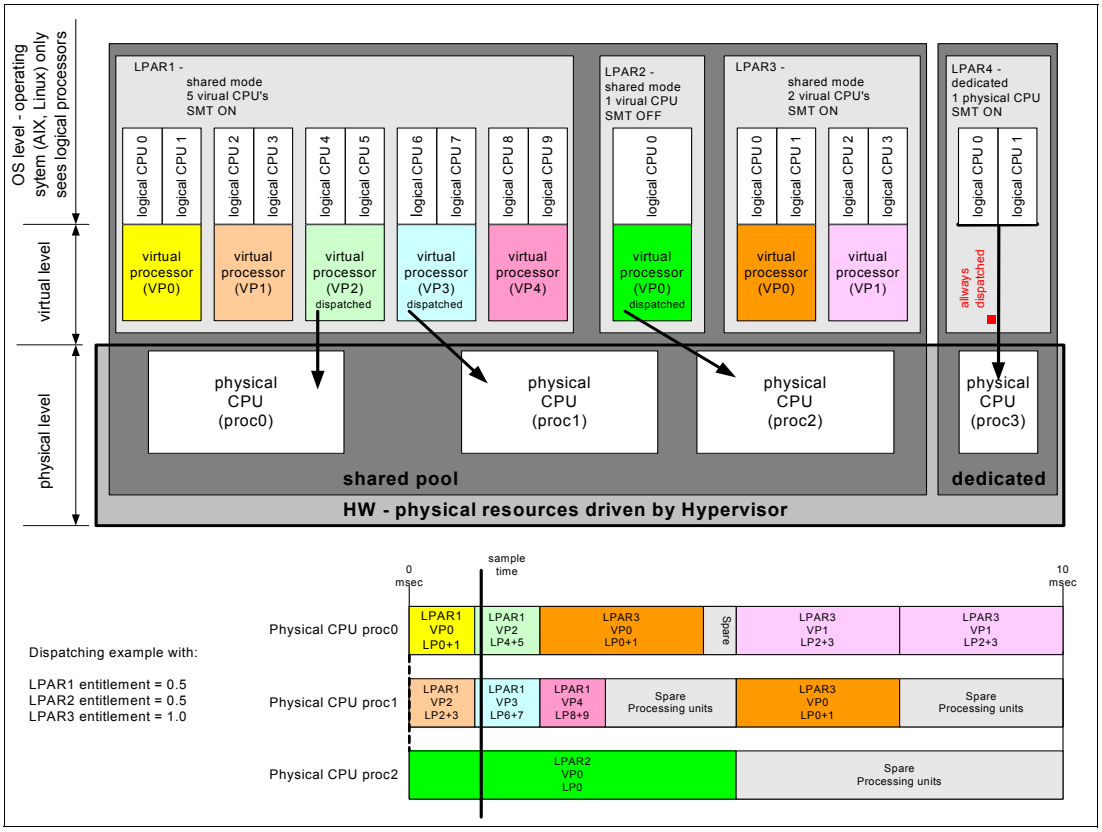


Figure 2-11 Logical, virtual, and physical processor mapping

In Figure 2-11, 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 thus three processors are available for shared processor pool. The LPAR1 defines five virtual processors and the simultaneous multithreading feature is on (thus 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 shared pool. Other virtual processors are idle waiting for dispatch by the Hypervisor. When more virtual processors are defined within a partition, any virtual processor share equal parts of partition processing entitlement.

2.12.3 Virtual I/O Server

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

Figure 2-12 on page 50 shows an organization view of micro-partitioned system including the Virtual I/O Server. The figure also includes virtual SCSI and Ethernet connections and mixed operating system partitions.

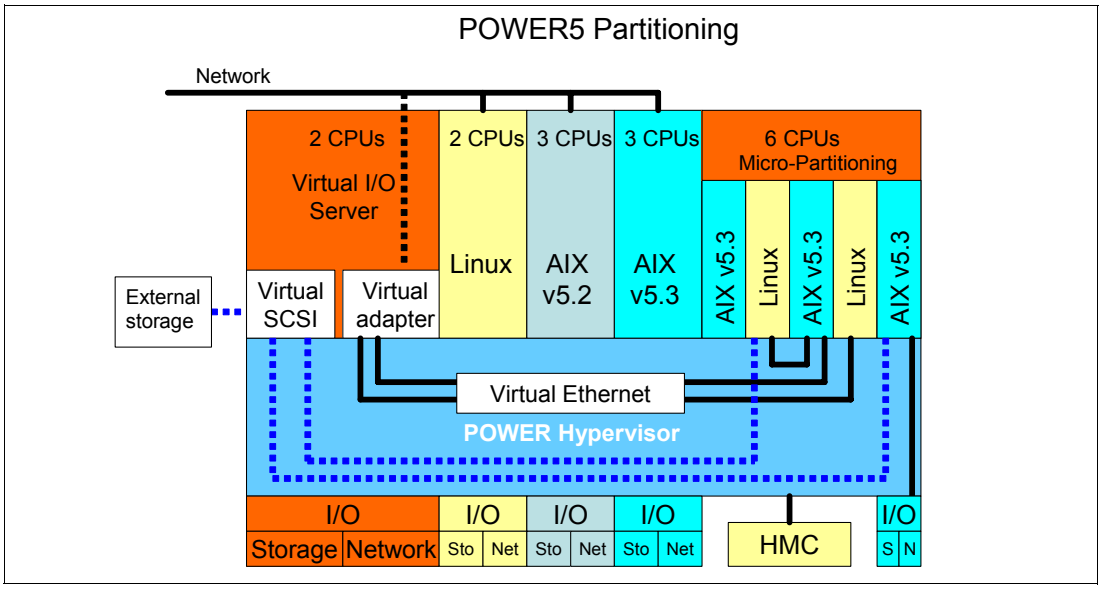


Figure 2-12 Micro-Partitioning technology and VIOS

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 by using 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 that order the Advanced POWER Virtualization feature. This dedicated software is only for the Virtual I/O Server (and IVM in case it is used) and is only supported in special Virtual I/O Server partitions.

The Virtual I/O Server can be installed by:

- Media (assigning the DVD-ROM drive to the partition and booting from the media)
- The HMC (inserting the media in the DVD-ROM drive on the HMC and using the **`installios`** command)
- Using 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 multipath driver). Logical volumes created on RAID or JBOD configurations are bootable, and the number of logical volumes is limited to the amount of storage available and 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 (SEA) 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 defined by Hypervisor on the Virtual I/O Server. A SEA enables LPARs on the virtual Ethernet to share access to the physical Ethernet and communicate with standalone 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 is 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 are required to 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 SEA setup can have up to 16 Virtual Ethernet trunk adapters and each Virtual Ethernet trunk adapter can support up to 20 VLAN networks. Therefore, it is possible for a single physical Ethernet to be shared between 320 internal VLAN. 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 as 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 using 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 using an HMC or through Integrated Virtualization Manager on smaller systems. The virtual SCSI server (target) adapter is responsible for executing any SCSI commands 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 and SAN attached devices and LUNs that are assigned to the client partition.

Physical disks owned by the Virtual I/O Server partition can either be exported and assigned to a client partition as 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 client partition point of view these two options are equivalent.

The Virtual I/O server provides mapping between *backing devices* (physical devices or logical volumes 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: Mirrored Logical Volumes (LVs) on Virtual I/O Server level are not recommended as backing devices. If mirroring is required, two independent devices (possibly from two separate VIO servers) should be assigned to the client partition and client partition should define mirror on top of them.

2.12.4 Partition Load Manager

Partition Load Manager (PLM) provides automated processor and memory distribution between a dynamic LPAR and a Micro-Partitioning technology capable logical partition running AIX 5L. The PLM application is based on a client/server model to share system information, such as processor or memory events, across 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: PLM is supported on AIX 5L Version 5.2 and AIX 5L Version 5.3, it is not supported on Linux.

2.12.5 Integrated Virtualization Manager

In order to ease virtualization technology adoption in any System p5 environment, IBM has developed Integrated Virtualization Manager (IVM), a simplified hardware management solution that inherits some HMC features, avoiding the necessity of a dedicated control workstation. This solution enables the administrator to reduce system setup time. IVM is targeted to small and medium systems; compared to HMC it has some important limitations listed in IVM limitations section on page 52.

The IVM provides a simple 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. Small and medium systems are best suited for the IVM.

IVM is an enhancement of Virtual I/O Server offered as part of Virtual I/O Server Version 1.2, 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.

IVM may 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 IVM managed server are as follows:

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

The major limitations of IVM in comparison to an HMC-managed system are as follows:

- All physical adapters are owned by IVM, and LPARs use virtual devices only. If IVM partition requires intervention for maintenance or other reason, all client partitions must be shut down as well, also virtual disks, optical, and Ethernet devices will not be accessible.
- No dynamic resource changes are allowed. Full dynamic LPAR support is for VIOS partition only, for example dynamically adding or removing memory or processing resources from a running client partition is not supported with the IVM. The partition should be powered off first. Keep in mind that the Hypervisor allows partitions to use more than their entitled processing capacity via the shared processing pool, lessening the importance of processing dynamic LPAR operations in some environments.
- Only one profile per partition.
- There are only four virtual Ethernet networks available inside the system.
- Each LPAR can have a maximum of one Virtual SCSI adapter assigned.
- It is not possible to have redundant Virtual I/O Servers because all I/O is managed by IVM.
- Service Agent (see 3.2.3, “Service Agent” on page 79) for reporting HW errors to IBM is not supported on IVM.
- IVM cannot be used by HACMP software to activate CoD resources on machines that support CoD.

Despite those limitations, IVM provides advanced virtualization functionality without the need for an extra-cost workstation. For more information about IVM 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 2-13 shows how a system with IVM is organized. There is a Virtual I/O server and IVM installed in one partition owning all physical server resources and four client partitions. IVM 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 2.12.2, “Logical, virtual, and physical processor mapping” on page 47 for shared partitions managed by the HMC.

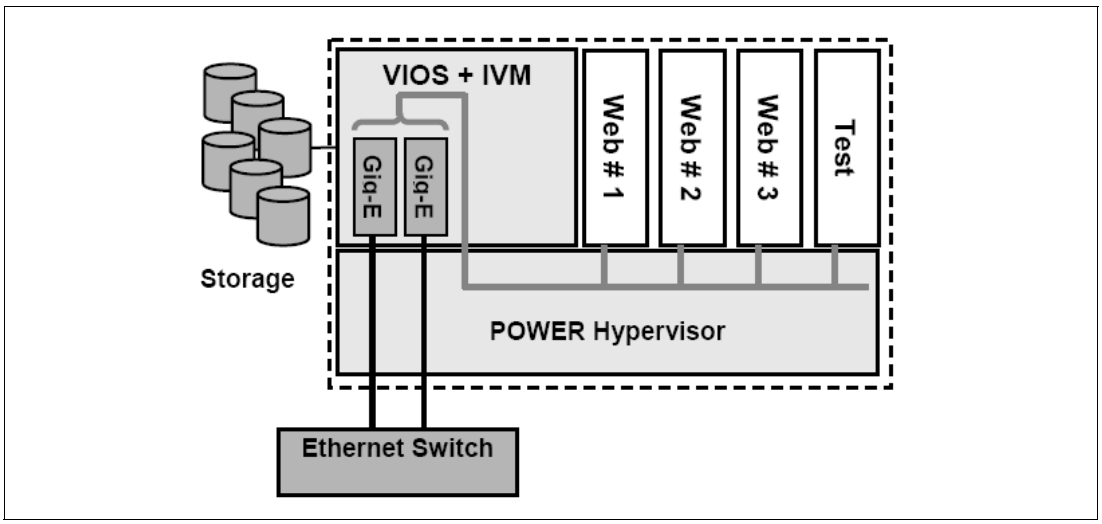


Figure 2-13 IVM principles

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

Operating system support for advanced virtualization

Table 2-10 lists AIX 5L and Linux support for advanced virtualization.

Table 2-10 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

2.13 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 functioning in either non-partitioned, LPAR, or clustered environments. In addition the Hardware Management Console 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:

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

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

- The HMC 7310 Model C04 is a desktop model with only one integrated 10/100/1000 Mbps Ethernet port, but two additional PCI slots.
- The HMC 7310 Model C05 is a deskside model with only one integrated 10/100/1000 Mbps Ethernet port, but 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 Virtual I/O Server for a unique connection from HMC to partitions. Therefore client partitions do not require own physical adapter in order to be able to communicate to HMC.

It is a good practise to connect the HMC to the first HMC Port on the server, 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 2-14 shows a simple network configuration to enable the connection from HMC to server, and to enable Dynamic LPAR operations. For more details about HMC and the possible network connections, refer to:

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

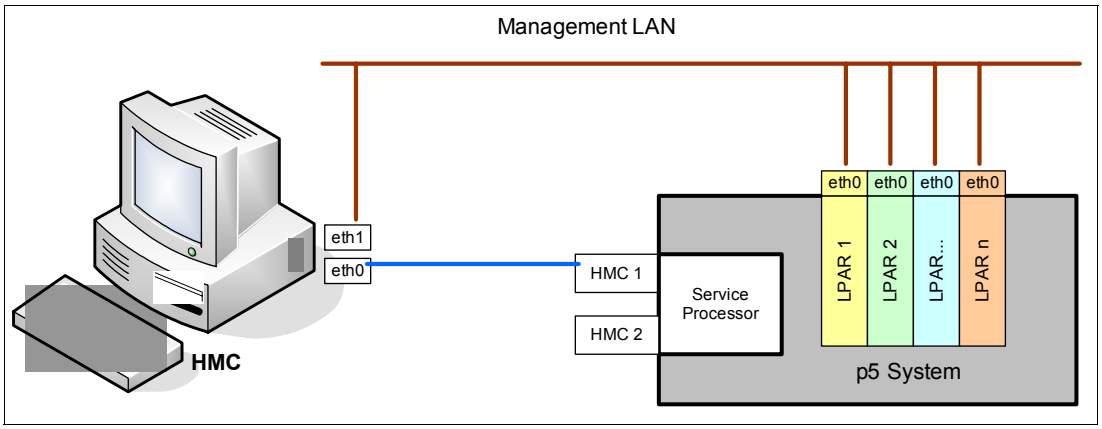


Figure 2-14 HMC to service processor and LPARs network connection

The default mechanism for 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 the managed server is powered on. If service processor of the managed server does not receive DHCP reply before time-out, predefined IP addresses will setup 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 by using the Advanced System Management Interface (ASMI) menus. See 2.15.7, “Service processor” on page 66 for predefined IP addresses and additional information.

Note: If you need to access ASMI (for example to setup 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

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 both graphical and command line interface for all management tasks. Remote connection to the HMC using Web-based System Manager or SSH are possible. For accessing the graphical interface, you can use the Web-based System Manager Remote Client running on UNIX (AIX 5L or 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 un-encrypted and encrypted Web-based System Manager connections are supported. The command line interface is also available by using 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.

2.13.1 High availability using the HMC

The HMC is an important hardware component. HAC MP Version 5.3 High Availability cluster software can be used to automatically activate resources (where available) thus becoming an integral part of the cluster. For some environments, it is recommended to work with redundant HMCs. POWER5 and POWER5+ processor-based systems have two service processor interfaces (HMC port 1 and HMC port 2) available for connections to the HMC. It is recommended to use both of them for redundant network configuration. Depending on your environment, you have multiple options to configure the network. Figure 2-15 shows one possible highly available configuration.

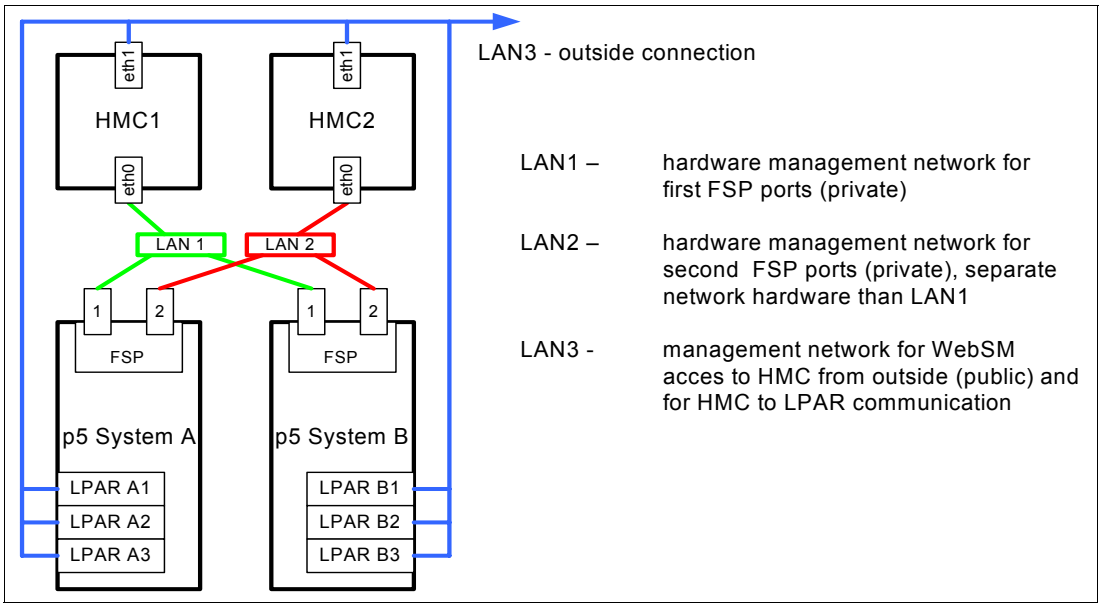


Figure 2-15 Highly available HMC and network architecture

Note that only hardware management network (LAN1 and LAN2) is highly available on the above picture in order to keep simplicity. But also management network (LAN3) can by made highly available by using similar concept and adding more Ethernet adapters to LPARs and HMCs.

2.13.2 LPAR validation tool

The IBM LPAR Validation Tool (LVT) is a tool available to assist the user in the design of LPARs and to provide an LPAR validation report that reflects the user's system requirements while not exceeding LPAR recommendations. The LVT is a PC based tool, standalone Java™ application, that runs on a Microsoft® Windows 95 or later workstation with 128 MB minimum of free memory and it is intended to run as a standalone Java application.

The LVT can be downloaded for free at:

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

The LVT provides a useful report that can complement the organization and validation of features required for the configuration of a complex partition solution. The LVT does support System p5 Express family and provides the following functions:

- Provides support for partitions running AIX 5L V5.2 and V5.3, and Linux.
- Validates dynamic LPAR design.
- Validates virtual partition design, including Virtual I/O Server and virtual clients.
- Calculates unallocated memory and shared processor pool.
- Calculates Hypervisor memory requirements.
- Calculates the number of operating system licenses needed to support partition design.
- Validates the number of virtual slots required for partitions.

Important: We recommend the use of the LVT to estimate Hypervisor requirements and determine memory resources required for all partitioned and non-partitioned servers.

Figure 2-16 shows the estimated Hypervisor memory requirements based on sample partition requirements.

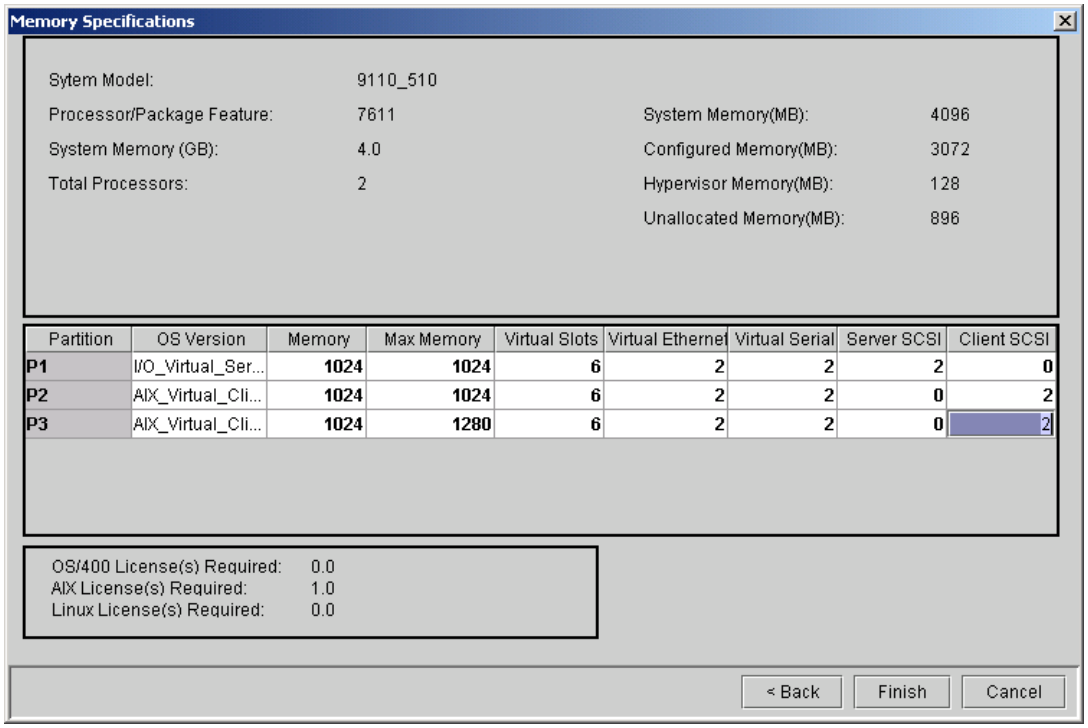


Figure 2-16 LVT window showing Hypervisor requirements

2.14 Operating system support

The p5-520 and p5-520Q are capable of running the AIX 5L and Linux operating systems. AIX 5L has been specifically developed and enhanced to exploit and support the extensive RAS features on IBM System p systems.

2.14.1 AIX 5L

If installing AIX 5L on the server, the following minimum requirements must be met:

- AIX 5L for POWER V5.2 with 5200-08 Technology Level (APAR IY77270), CD# LCD4-1133-08 or later, DVD LCD4-7549-01 media is also available.
- AIX 5L for POWER V5.3 with 5300-04 Technology Level (APAR IY77273), CD#LCD4-7463-05 or later, DVD LCD4-7544-01 media is also available.

Note: The Advanced POWER Virtualization feature (FC 7940) 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 they can be downloaded from the Internet at:

`http://www.ibm.com/servers/eserver/support/pseries/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 site. In AIX 5L V5.3, the **suma** command is also available, which helps the administrator to automate the task of checking and downloading operating system downloads. For more information about the **suma** command, refer to:

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

On November 18, 2005, 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, opposite to waiting for a physical shipment to arrive. Clients requesting ESD should order FC 3450.

ESD has the following requirements:

- POWER5 system
- Internet connectivity from a POWER5 system or PC, reasonable connection speed for downloading large products such as AIX 5L
- Registration on the ESD Web site

For additional information, contact your IBM sales representative.

Software support for new features in POWER5+ processor

For complete list of new features introduced in POWER5+ processor, see 2.1, “The POWER5+ chip” on page 24. Support for two new virtual memory page sizes was introduced - 64 KB and 16 GB as well as support for 1 TB segment size. While 16 GB pages are intended to only be used in very high performance environments, 64 KB pages are general-purpose. AIX 5L Version 5.3 with the 5300-04 Technology Level 64-bit kernel is required for 64 KB and 16 GB page size support.

As with all previous versions of AIX, 4 KB is the default page size. A process will continue to use 4 KB pages unless a user specifically requests another page size be used. AIX 5L has rich support of 64 KB pages. They are easy to use, and it is expected that many applications will see performance benefits when using 64 KB pages rather than 4 KB pages. No system configuration changes are necessary to enable a system to use 64 KB pages, they are fully pageable, and the size of the pool of 64 KB page frames on a system is dynamic and fully managed by AIX 5L.

The main benefit of a larger page size is improved performance for applications that allocate and repeatedly access large amounts of memory. The performance improvement from larger page sizes is due to the overhead of translating a page address as it is used in an application, to a page address that is understood by the computer's memory subsystem. To improve performance, the information needed to translate a given page is usually cached in the processor. In POWER5+, this cache takes the form of a translation lookaside buffer (TLB). Since there are a limited number of TLB entries, using a large page size increases the amount of address space that can be accessed without incurring translation delays. Also the size of TLB in POWER5+ has been doubled compared to POWER5

Huge pages (16 GB) are intended to be used only in very high performance environments, and AIX 5L will not automatically configure a system to use these page sizes. A system administrator must configure AIX 5L to use these page sizes and specify their number via HMC before partition start.

A user may specify page sizes to use for three regions process's address space with an environment variable or with settings in an application's XCOFF binary using the **ldedit** or **ld** commands. These three regions are: data, stack and program text. An application programmer can also select the page size to use for System V shared memory via a new SHM_PAGESIZE command to the shmctl() system call.

An example of using system variables to start a program with 64 KB page size support:

```
LDR_CNTRL=DATAPSIZE=64K@TEXTPSIZE=64K@STACKPSIZE=64K <program>
```

Systems commands (ps, vmstat, svmon, pagesize) have been enhanced to report various page size usage.

2.14.2 Linux

For the p5-520 and p5-520Q, Linux distributions are available through Novel SUSE and Red Hat at the time this publication was written. The server requires the following version of Linux distributions:

- SUSE Linux Enterprise Server 9 for IBM POWER Service Pack 3, or later
- Red Hat Enterprise Linux AS 4 for IBM POWER Service Update 2, or later

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

For information about the features and external devices supported by Linux, 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 dependant and might not be available on Linux. For more information, see:

http://www.ibm.com/servers/eserver/linux/power/whitepapers/linux_overview.html

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

Specially priced Linux subscriptions

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

These new Linux options, available on System p5 Express 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 Edition 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 Edition, 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 on POWER, visit:

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

2.15 Service information

The p5-520 and p5-520Q are a customer setup server (CSU) and is shipped with materials to assist to 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.

2.15.1 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, installing additional processors contained in the second processor book.

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 has to be performed first. One example is deconfiguring a physical volume in the operating system before removing a disk from a 4-pack disk enclosure of the p5-520 and p6-520Q.

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.

2.15.2 Securing a rack-mounted system into a rack

The *optional* rack-mount drawer rail kit is a unique kit designed for use with the rack-mounted model. No tools are required to install the server 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.

Note: Always exercise standard safety precautions when installing or removing devices from racks. By placing the rack-mounted system or expansion unit in the service position, you can access the inside of the unit.

2.15.3 Serving a rack-mounted system into a rack

To place the rack-mounted system or expansion unit into the service position, follow these steps:

1. If necessary, open the front rack door.
2. Remove the two thumbscrews A that secure the system or expansion unit B to the rack as shown in the Figure 2-17.

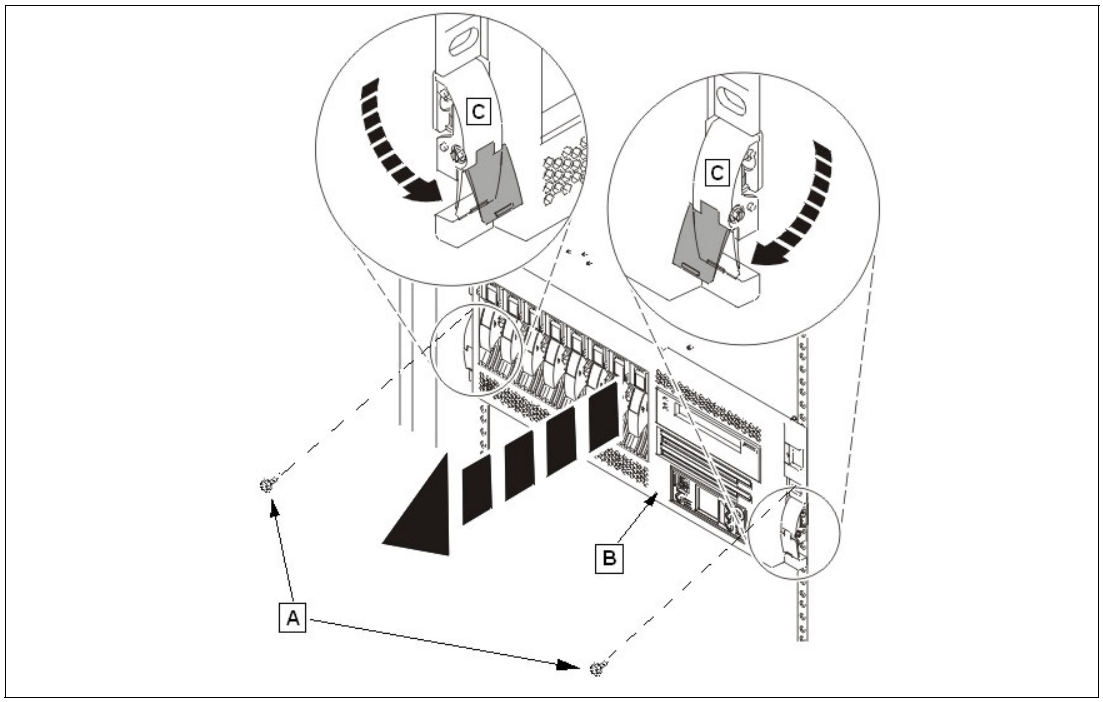


Figure 2-17 Pull the server to the service position

3. Release the rack latches C on both the left and right sides as shown in the Figure 2-17.
4. Review the following notes, and then slowly pull the system or expansion unit out from the rack until the rails are fully extended and locked.
   - If the procedure you are performing requires you to unplug cables from the back of the system or expansion unit, do so before you pull the unit out from the rack.
   - Ensure that the cables at the rear of the system or expansion unit do not catch or bind as you pull the unit out from the rack.

– When the rails are fully extended, the rail safety latches lock into place. This action prevents the system or expansion unit from being pulled out too far.

Caution: This unit weighs approximately 43 kg (95 lb.). Ensure that you can safely support this weight when removing the server unit from the system rack.

The *IBM Systems Hardware Information Center* is available for more information, or to view available video-clips that describes several of the maintenance repair-action procedures.

2.15.4 Cable-management arm

The rack-mounted model is shipped with a cable-management arm to route all the cables through the hooks along the cable arm and secure them with the straps provided. The cable-management arm simplify the cables management in case of service action that require to pull-out the rack-mounted system from the rack.

2.15.5 Operator control panel

The service processor provides an interface to the control panel that is used to display server status and diagnostic information. See Figure 2-18 for operator control panel physical details and buttons.

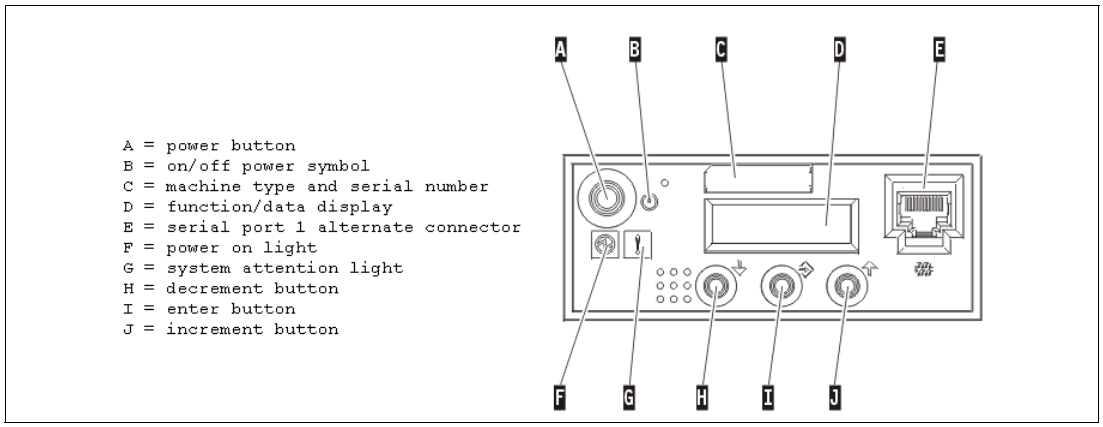


Figure 2-18 Operator control panel physical details and button

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

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

Extended control panel functions

The extended control panel functions consist of two major groups:

- Functions 21 through 49, which are available when you select Manual mode from Function 02.
- Support service representative Functions 50 through 99, which are available when you select Manual mode from Function 02, then select and enter the customer service switch 1 (Function 25), followed by service switch 2 (Function 26).

Function 30 – CEC SP IP address and location

Function 30 is one of the Extended control panel functions and is only available when Manual mode is selected. This function can be used to display the central electronic complex (CEC) Service Processor IP address and location segment. The Table 2-11 shows an example of how to use the Function 30:

Table 2-11 CEC SP IP address and location

Information on operator panel	Action or description
3 0	Use the increment or decrement buttons to scroll to Function 30
3 0 * *	Press Enter button to enter sub-function mode
3 0 0 0	Use the increment or decrement buttons to select an IP address: 0 0 = Service Processor ETH0 or HMC1 port 0 1 = Service Processor ETH1 or HMC2 port
S P A: E T H 0: _ _ _ T 5 1 9 2 . 1 6 8 . 2 . 1 4 7	Press Enter button to display the selected IP address
3 0 * *	Use the increment or decrement buttons to select sub-function exit
3 0	Press Enter to exit sub-function mode

2.15.6 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. See “IBM System p5 firmware maintenance” on page 81 for a detailed description of system p5 firmware.

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 will boot, 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.

System firmware download site

For the system firmware download site, go to:

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

Receive server firmware fixes using an HMC

If you use an HMC to manage your server and you periodically need to 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 two 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.

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.

Install the server firmware fix permanently (optional).

2.15.7 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.

Service processor is used to monitor and manage the system hardware resources and devices. The service processor offers the following connections:

- 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 major function of Service Processor, see 3.2.1, “Service processor” on page 77.

2.15.8 Hardware management user interfaces

This section provides a brief overview of the different 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 directly to the service processor (in this case a standard Ethernet cable or a crossed cable can be both used) or through an Ethernet network. Using the *network configuration menu*, the ASMI enables the possibility to change the service processor IP addresses or to apply some security policies and avoid the access from undesired IP addresses or range. The ASMI can also be accessed using a terminal attached to the system service processor ports on the server, if the server is not HMC managed. The service processor and the ASMI are standard on all IBM System p 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, Mozilla Firefox, 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 with.
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 changing the boot list, or 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 POST[3] indicators memory, keyboard, network, SCSI, 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.

The SMS menus is useful to define the operating system installation method, choosing the installation boot device or setting the boot device priority list for a full managed server or a logical partition. In case of a network boot, SMS menus provide to setup the network parameters and network adapter IP address.

HMC

The Hardware Management Console is a system that controls managed systems, including IBM System p5 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 can be a DHCP[4] server from

[3] POST stands for power-on-self-test.
[4] DHCP stands for Dynamic Host Control Protocol.

which the managed server obtains the address for its firmware. Additional local HMCs in your private service network cannot be other DHCP server but they can be 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 and can be downloaded directly from an installed HMC. Once a HMC is installed and HMC Ethernet IP addresses have been assigned, it is possible to download the Web-based System Manager Remote Client from a web browser, using the following URL:

```
http://HMC_IP_address/remote_client.html
```

You can then use the 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 using the ASMI and selecting the boot to Open Firmware prompt.

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

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

3

RAS and manageability

The following sections provide more detailed information about IBM System p5 design features that will help lower the total cost of ownership (TCO). IBM RAS (Reliability, Availability, and Service ability) technology allows possibility to improve your TCO architecture by reducing unplanned down time. This chapter includes several features based on the benefits available when using AIX 5L. Support of these features using Linux can vary.

3.1 Reliability, Availability and Serviceability

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 provided by both the hardware and the operating system AIX 5L.

The p5-520 or p5-520Q as a POWER5+ server enhances the RAS capabilities 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

System p5 servers are built on a quality-based design intendedkeep errors from happening. This design includes the following features:

- Reduced power consumption, cooler operating temperatures for increased reliability, enabled by the use of copper chip 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 correctly diagnose it is a fundamental requirement upon which improved availability is based. The p5-520 and p5-520Q incorporate advanced capability in start-up diagnostics and in run-time First Failure Data Capture (FDDC) based on strategic error checkers built into the chips.

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 73 shows a schematic of a Fault Register Implementation.

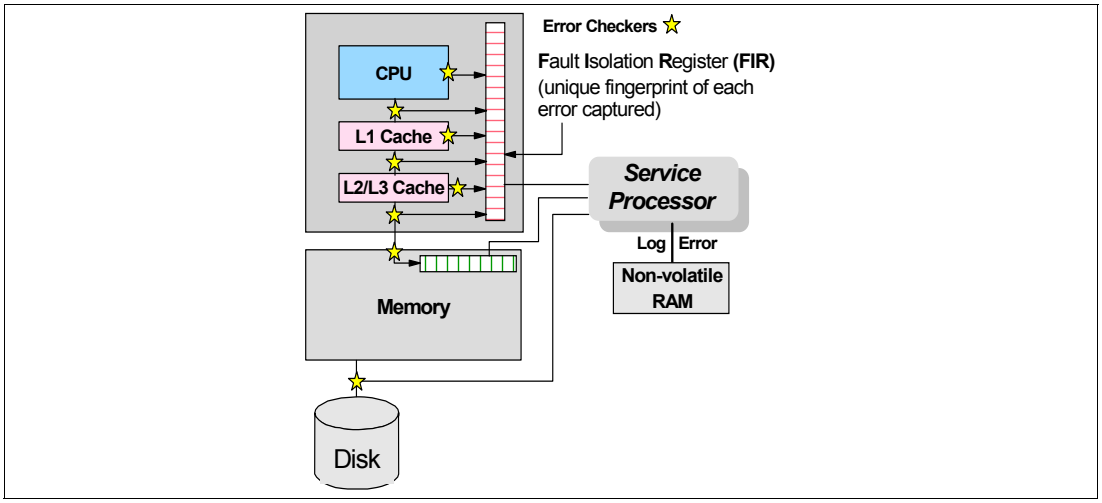


Figure 3-1 Schematic of Fault Isolation Register implementation

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 accurately record any future errors.

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 75.

3.1.3 Permanent monitoring

The SP included in the p5-520 or p5-520Q provides a way to monitor the system even when the main processor is inoperable. See the next subsection for a more detailed description of monitoring functions in p5-520 and p5-520Q.

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 so as to prevent any data

loss without the 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 fans 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 module to keep the server operational
- *Bit-scattering*, thus allowing for error correction and continued operation in the presence of a complete chip failure (Chipkill™ recovery)
- Single bit error correction using 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 p5-520 and p5-520Q 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 module 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 modules 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 chip and does not influence normal system memory performance.

3.1.5 N+1 redundancy

The use of redundant parts allows the p5-520 and p5-520Q to remain operational with full resources:

- Redundant spare memory bits in L1, L2, L3, and main memory

- Redundant fans
- Redundant power supplies (optional)

Note: With this optional feature every deskside or rack-mount p5-520 or p5-520Q, requires two power cords, which are not included in the base order. For maximum availability it is highly recommended to connect power cords from same p5-520 or p5-520Q to two separate PDUs in the rack. These PDUs being connected to two independent client power sources. For deskside p5-520 or p5-520Q power cords need to be plugged to two independent power source in order to achieve maximum availability.

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 non disruptive, 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 de-configuration 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 chip 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 will result in recovery action. If this is unsuccessful, the PCI device will be gracefully shut down. Parity errors on the PCI bus itself will result in bus retry, and if uncorrected, the bus and any I/O adapters or devices on that bus will be de-configured.

The p5520 or p5-520Q 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 p5-520 or p5-520Q 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 will examine the affected bus, allow the device driver to reset it, and continue 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 will invoke extended diagnostics. If a processor or L3 cache has been marked for de-configuration by persistent processor deallocation, the boot process will attempt to proceed to completion with the faulty device automatically de-configured. Failing I/O adapters will be de-configured 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. p5-520 and p5-520Q improve service productivity by providing the functions described in the following subsections:

Error indication and LED indicators

The p5-520 and p5-520Q are designed for client setup of the machine and for the subsequent addition of most hardware features. The p5-520 and p5-520Q also allow clients to replace service parts (Client Replaceable Unit). To accomplish this, the p5-520 or p5-520Q provide internal LED diagnostics that will 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/eserver

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 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
- Operator Panel (requires HMC guided support)
- GX RIO-2/HSL-2 Adapter (FC 2888).
  - All PCI-X adapters connected to the involved RIO loop must be first varied offline from the operating system.
  - This concurrent maintenance task requires HMC guided support.

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 regardless of main p5 Central Electronic Complex (CEC) state. 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 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 will not boot. It can be performed either form ASMI, HMC or ASCI console (depending on presence of HMC). ASMI is provided in any case.

See Figure 3-2 for an example of the ASMI accessed from a Web browser.

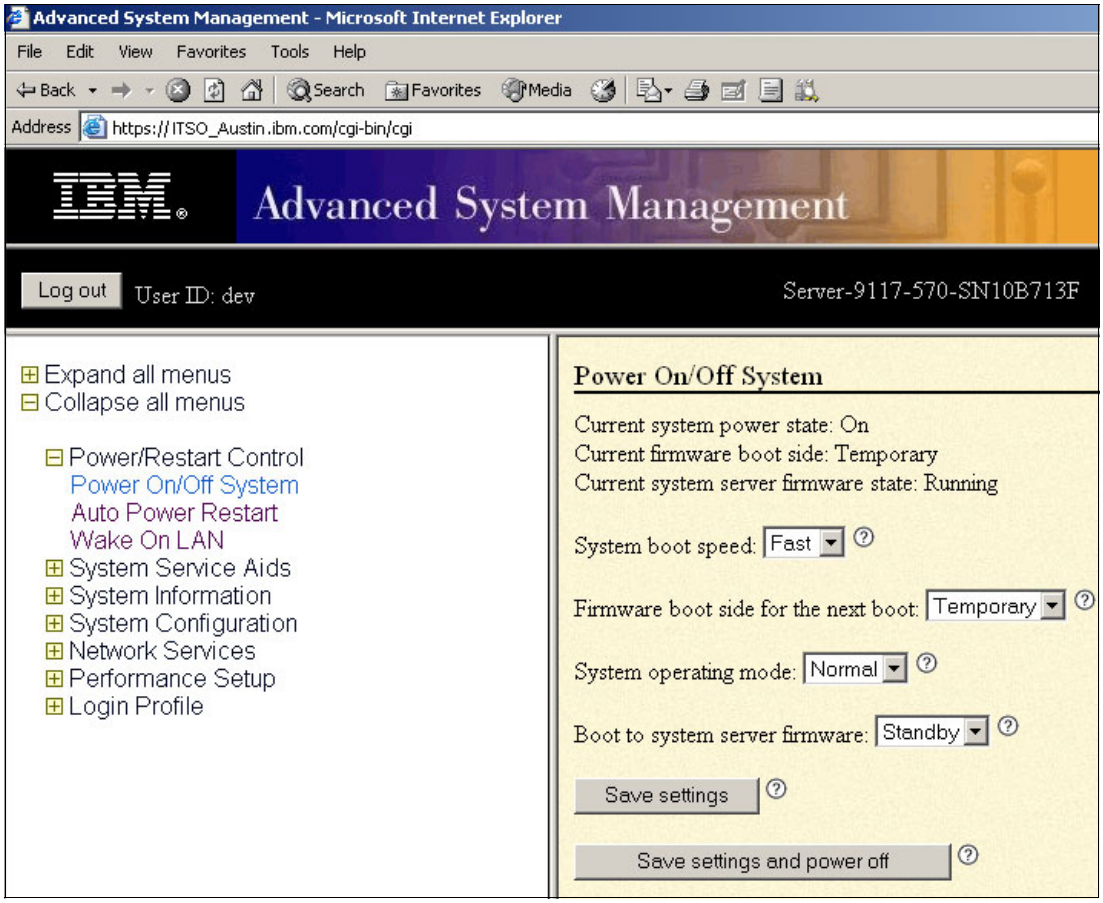


Figure 3-2 Advanced System Management main menu

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 may 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 be 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 p5-520 and p5-520Q system have an optional DVD-ROM (FC 1994) and DVD-RAM (FC 1993), 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 them 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 System p5, pSeries, and RS/6000 include:

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

In addition:

- 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. Includes high-availability cluster workstation (HACWS) for 9076.
- Network environment support with minimum telephone lines for modems.
- Provides communication base 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

Machines are defined by using the Service Agent user interface. After the machines are defined, they are registered with the IBM Service Agent Server (SAS). During the registration process, an electronic key is created that becomes part of your resident Service Agent program. This key is used 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 service call is refused and posted back using an e-mail notification.

Service agent can be configured to connect to IBM either using modem or 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, 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 how can an IBM System p5 system be configured to utilize all features of Service Agent. The shown communication to IBM support can be either modem or network. If HMC is present, Service Agent is an integral part of it and if activated will collect hardware related information a error messages about the whole system and partitions. If software level information (like performance data for example) is also required, Service Agent can also be installed on any of the partitions and can be configured to act as either gateway and connection manager or a client. Gateway and connection manager gathers data from clients and communicates to IBM on behalf of them.

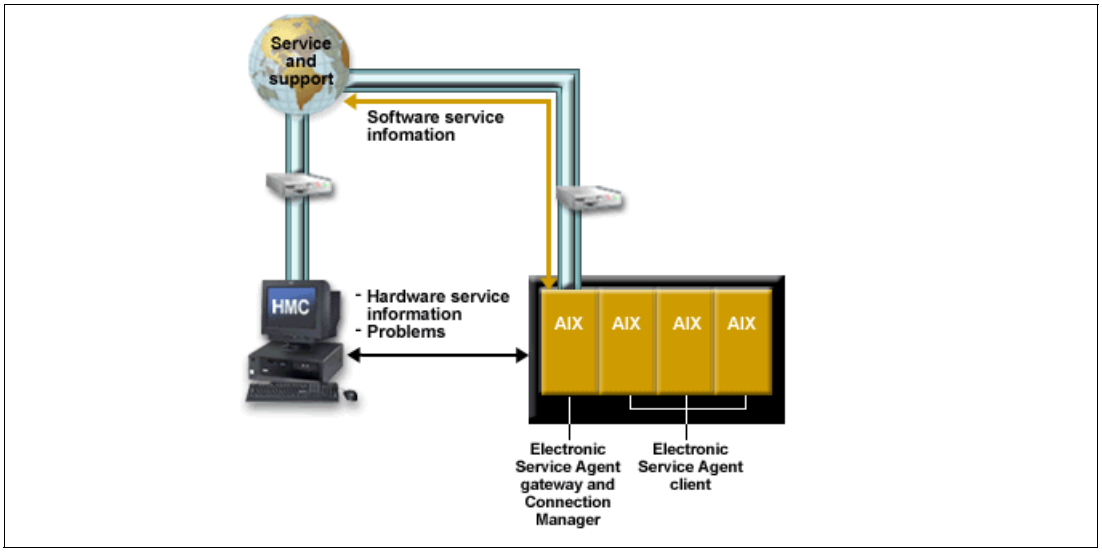


Figure 3-3 Service agent and possible connections to IBM

Additional services provided by Service Agent:

- My Systems: Client and IBM persons authorized by client can view HW information and error messages gathered by Service Agent on Electronic Services WWW pages (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 means for collecting long term performance data. The data are collected in reports accessed by 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) but 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 System p5, pSeries, and RS/6000 systems and associated I/O adapters. The IBM System p5 microcode can be installed either from HMC or from a running partition, for update details, see 2.15.6, “System firmware” on page 63.

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 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 “t” side of system firmware, for additional discussion about firmware sides see 2.15.7, “Service processor” on page 66.
- 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 “p” side of system firmware, for additional discussion about firmware sides see 2.15.7, “Service processor” on page 66.

IBM introduced the Concurrent Firmware Maintenance (CFM) function on System p5 systems in system firmware level 01SF230_126_120, which was released on June 16, 2005. This function supports non disruptive 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 may, on occasion, contain fixes that are known as deferred. These deferred fixes can be installed concurrently, but will not be activated until the next IPL. Deferred fixes, if any, will be identified in the Firmware Update Descriptions table of

this document. For deferred fixes within a service pack, only the fixes in the service pack which cannot be concurrently activated are deferred.

Use the following information as a reference to determine whether your installation will be concurrent or disruptive.

Figure 3-4 shows the system firmware file naming convention:

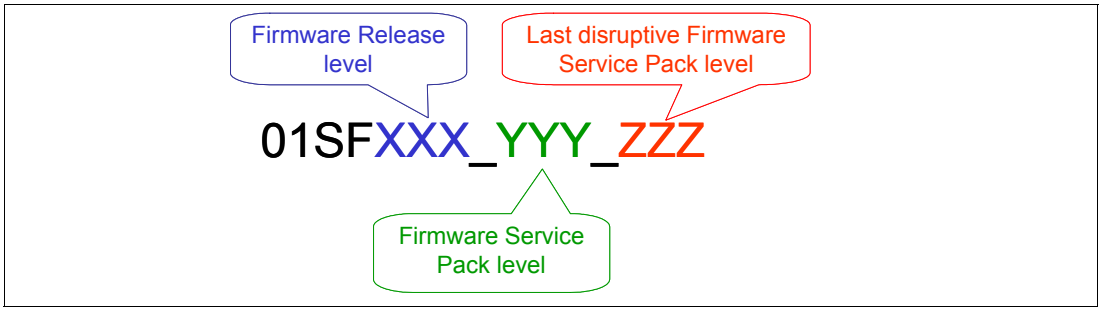


Figure 3-4 System firmware file naming convention

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 new firmware.

Otherwise an installation is concurrent if:

- If the service pack level (YYY) of new the 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 downtime
- 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 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 p5-520 and p5-520Q. For hardware control, an HMC is required. One HMC can also control several 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/servers/eserver/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 automatically installed and ready when you install AIX 5L, so each system or logical partition is cluster-ready.

The CSM V1.4 on AIX 5L and Linux introduces an optional IBM CSM High Availability Management Server (HA MS) 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 NTP[1], 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.

[1] Network Time Protocol

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 87. Note that some of the documents referenced here may be available in softcopy only.

- *IBM @server p5 590 and 595 System Handbook*, SG24-9119
- *Partitioning Implementations for IBM @server p5 Servers*, SG24-7039
- *Managing AIX Server Farms*, SG24-6606
- *Practical Guide for SAN with pSeries*, SG24-6050
- *Problem Solving and Troubleshooting in AIX 5L*, SG24-5496
- *Understanding IBM @server pSeries Performance and Sizing*, SG24-4810
- *Advanced POWER Virtualization on IBM System p5*, SG24-7940
- *Virtual I/O Server Integrated Virtualization Manager*, REDP-4061
- *IBM IntelliStation POWER 185 Technical Overview and Introduction*, REDP-4135
- *IBM System p5 185 Technical Overview and Introduction*, REDP-4141
- *IBM System p5 505 Express Technical Overview and Introduction*, REDP-4079
- *IBM System p5 510 and 510QTechnical Overview and Introduction*, REDP-4136
- *IBM System p5 550 and 550Q Technical Overview and Introduction*, REDP-4138
- *IBM System p5 560Q Technical Overview and Introduction*, REDP-4139
- *IBM System p5 Quad-Core Module Based On POWER5+ Technology Technical Overview and Introduction*, REDP-4150
- *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 @server p5 590 and 595 Technical Overview and Introduction*, REDP-4024

Other publications

These publications are also relevant as further information sources:

- *7014 Series Model T00 and T42 System Rack Service Guide*, SA38-0577, contains information regarding the 7014 Model T00 and T42 Racks, 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.

- *RS/6000 and @server pSeries Adapters, Devices, and Cable Information for Multiple Bus Systems*, SA38-0516, contains information about adapters, devices, and cables for your system.
- *RS/6000 and @server pSeries PCI Adapter Placement Reference*, 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.
- *IBM System p5, @server p5 and i5, and OpenPower Planning*, SA38-0508, contains site and planning information, including power and environment specification.

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/pseries/aixfixes.html
- IBM System p5, @server p5, pSeries, OpenPower, and IBM RS/6000 Performance Report

  http://www.ibm.com/servers/eserver/pseries/hardware/system_perf.html
- IBM TotalStorage Expandable Storage Plus

  http://www.ibm.com/servers/storage/disk/expplus/index.html
- IBM TotalStorage Mid-range Disk Systems

  http://www.ibm.com/servers/storage/disk/ds4000/index.html
- IBM TotalStorage Enterprise disk storage

  http://www.ibm.com/servers/storage/disk/enterprise/ds_family.html
- IBM Virtualization Engine

  http://www.ibm.com/servers/eserver/about/virtualization/
- Advanced POWER Virtualization on IBM @server p5

  http://www.ibm.com/servers/eserver/pseries/ondemand/ve/resources.html
- Virtual I/O Server supported environments

  http://www14.software.ibm.com/webapp/set2/sas/f/vios/home.html
- Hardware Management Console support information

  http://techsupport.services.ibm.com/server/hmc
- IBM LPAR Validation Tool (LVT), a PC-based tool intended assist you in logical partitioning

  http://www.ibm.com/servers/eserver/iseries/lpar/systemdesign.htm
- Customer Specified Placement and LPAR Delivery

  http://www.ibm.com/servers/eserver/power/csp/index.html
- SUMA on AIX 5L

  http://techsupport.services.ibm.com/server/suma/home.html
- Linux on IBM @server p5 and pSeries

  http://www.ibm.com/servers/eserver/pseries/linux/
- SUSE Linux Enterprise Server 9

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

- Red Hat Enterprise Linux details

  http://www.redhat.com/software/rhel/details/
- IBM @server Linux on POWER Overview

  http://www.ibm.com/servers/eserver/linux/power/whitepapers/linux_overview.html
- Autonomic computing on IBM @server pSeries servers

  http://www.ibm.com/autonomic/index.shtml
- *IBM @server p5 AIX 5L Support for Micro-Partitioning and Simultaneous Multi-threading* whitepaper

  http://www.ibm.com/servers/aix/whitepapers/aix_support.pdf
- Hardware documentation

  http://publib16.boulder.ibm.com/pseries/en_US/infocenter/base/
- IBM Systems Information Center

  http://publib.boulder.ibm.com/eserver/
- IBM @server pSeries support

  http://www.ibm.com/servers/eserver/support/pseries/index.html
- IBM @server support: Tips for AIX 5L administrators

  http://techsupport.services.ibm.com/server/aix.srchBroker
- Linux for IBM @server pSeries

  http://www.ibm.com/servers/eserver/pseries/linux/
- Microcode Discovery Service

  http://techsupport.services.ibm.com/server/aix.invscoutMDS
- POWER4 system microarchitecture, comprehensively described in the *IBM Journal of Research and Development*, Vol 46, No.1, January 2002

  http://www.research.ibm.com/journal/rd46-1.html
- SCSI T10 Technical Committee

  http://www.t10.org
- Microcode downloads for IBM @server i5, OpenPower, p5, pSeries, and RS/6000 Systems

  http://techsupport.services.ibm.com/server/mdownload
- Resource Link

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

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 520 and 520Q Technical Overview and Introduction

Redpaper

Finer system granulation using Micro-Partitioning technology to help lower TCO

Support for versions of AIX 5L and Linux operating systems

From Web servers to integrated cluster solutions

This document is a comprehensive guide covering the IBM® System p5™ 520 and 520Q UNIX® servers. It introduces major hardware offerings and discusses their prominent functions. Professionals wishing to acquire a better understanding of IBM System p products should read this document. The intended audience includes:

Clients
Sales and marketing professionals
Technical support professionals
IBM Business Partners
Independent software vendors

This document expands the current set of IBM System p and documentation by providing a desktop reference that offers a detailed technical description of the p5-520 and the p5-520Q system.

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

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