IBM

IBM Power 520 Technical Overview

Covering the 8203-E4A supporting the AIX, i, and Linux
operating systems

POWER6 processor-based PowerVM
virtualization technology

Delivers IBM POWER6
efficiency
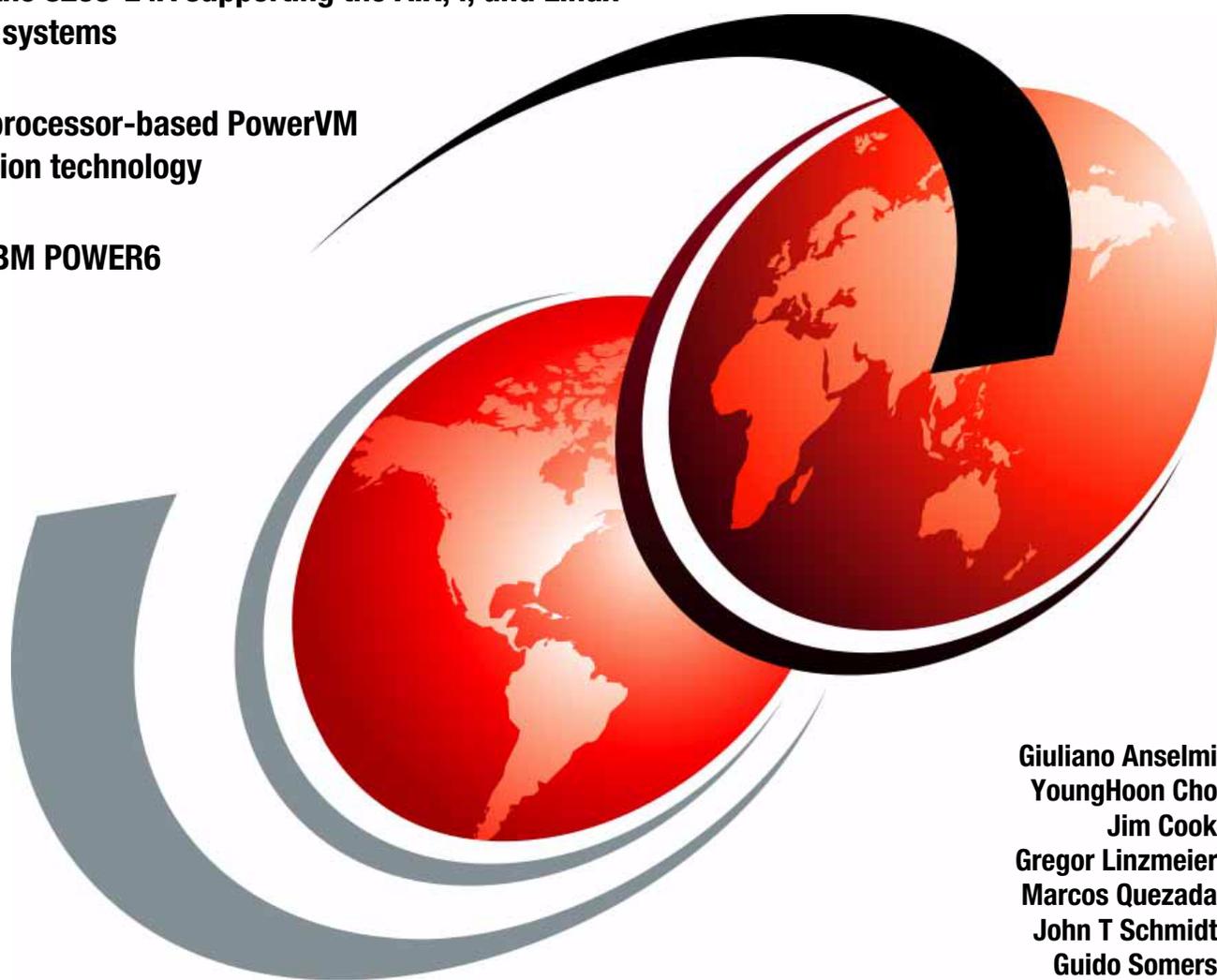
Giuliano Anselmi
YoungHoon Cho
Jim Cook
Gregor Linzmeier
Marcos Quezada
John T Schmidt
Guido Somers

Redpaper

IBM

International Technical Support Organization

IBM Power 520 Technical Overview

May 2009

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

First Edition (May 2009)

This edition applies to the IBM Power 520 8203-E4A server. The following IBM operating systems are included: AIX 5.3 (product number 5765-G53), AIX 6.1 (product number 5765-G62), IBM i operating system releases 5.4 (5722-SS1) and 6.1 (5761-SS1).

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 illustrate 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.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. These and other IBM trademarked terms are marked on their first occurrence in this information with the appropriate symbol (® or ™), indicating US registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at http://www.ibm.com/legal/copytrade.shtml

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

1350™	IBM®	pSeries®
Active Memory™	iCluster®	Rational®
AIX 5L™	IntelliStation®	Redbooks®
AIX®	iSeries®	Redbooks (logo) ®
BladeCenter®	LoadLeveler®	RS/6000®
DB2®	Micro-Partitioning™	Solid®
DS4000®	OpenPower®	System i5®
DS6000™	OS/400®	System i®
DS8000®	Power Architecture®	System p5®
Electronic Service Agent™	POWER Hypervisor™	System p®
EnergyScale™	Power Systems™	System Storage™
Enterprise Storage Server®	POWER4™	System x®
eServer™	POWER5™	System z®
Express™	POWER5+™	Tivoli®
FlashCopy®	POWER6™	TotalStorage®
Focal Point™	POWER6+™	VisualAge®
GPFS™	PowerExecutive™	WebSphere®
HACMP™	PowerHA™	Workload Partitions Manager™
i5/OS®	PowerPC®	XIV®
IBM Systems Director Active Energy Manager™	PowerVM™ POWER®	z/OS®

The following terms are trademarks of other companies:

AMD, AMD Opteron, the AMD Arrow logo, and combinations thereof, are trademarks of Advanced Micro Devices, Inc.

BNT, and the BLADE logo are registered trademarks of BLADE Network Technologies, Inc.

InfiniBand Trade Association, InfiniBand, and the InfiniBand design marks are trademarks and/or service marks of the InfiniBand Trade Association.

Data ONTAP, and the NetApp logo are trademarks or registered trademarks of NetApp, Inc. in the U.S. and other countries.

Novell, SUSE, the Novell logo, and the N logo are registered trademarks of Novell, Inc. in the United States and other countries.

Oracle, JD Edwards, PeopleSoft, Siebel, and TopLink are registered trademarks of Oracle Corporation and/or its affiliates.

QLogic, and the QLogic logo are registered trademarks of QLogic Corporation. SANblade is a registered trademark in the United States.

Red Hat, and the Shadowman logo are trademarks or registered trademarks of Red Hat, Inc. in the U.S. and other countries.

ABAP, mySAP, SAP NetWeaver, SAP, and SAP logos are trademarks or registered trademarks of SAP AG in Germany and in several other countries.

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

AltiVec is a registered trademark of Motorola.

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

Intel Xeon, Intel, Intel logo, Intel Inside logo, and Intel Centrino logo are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States, other countries, or both.

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

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

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

Preface

This IBM® Redpaper publication is a comprehensive guide covering the IBM Power System 520 server. The goal of this paper is to provide a hardware-based overview of the POWER6™ processor-based Power 520 server configurations that encompass System i® and System p® capabilities and features and introduce POWER6 capabilities.

We also associate marketing announcement terminology with more specific machine type model (MTM) terminology that includes common features and those unique and familiar to System i and System p customers.

IBM System i and IBM System p platforms are unifying the value of their servers into a single, powerful portfolio of servers based on industry-leading POWER6 processor technology with support for the IBM i operating system (formerly known as i5/OS®), and IBM AIX® and Linux® operating systems.

The Power 520 model is the entry member of the portfolio. Key hardware-oriented capabilities provided by the Power 520 include:

► The POWER6 processor technology available at a frequency of 4.2 GHz and the POWER6+™ processor available at a frequency of 4.7 GHz.

► The specialized POWER6 DDR2 memory provides greater bandwidth, capacity, and reliability.

► The 1 Gb or 10 Gb Integrated Virtual Ethernet adapter, included with each server configuration, provides native hardware virtualization.

► EnergyScale™ technology that provides features such as power trending, power-saving, capping of power, and thermal measurement.

► PowerVM™ Virtualization.

► Mainframe level continuous availability capabilities.

Professionals wishing to acquire a better understanding of POWER6 products should read this Redpaper. The intended audience includes:

► Clients

► Sales and marketing professionals

► Technical support professionals

► IBM Business Partners

► Independent software vendors

This Redpaper expands the current set of IBM Power Systems™ documentation by providing a desktop reference that offers a detailed technical description of the Power 520 server.

This Redpaper does not replace the latest marketing materials, tools, and other IBM publications available, for example, at the IBM Systems Hardware Information Center:

   http://publib.boulder.ibm.com/infocenter/systems/scope/hw/index.jsp

It is intended as an additional source of information that, together with existing sources, can be used to enhance your knowledge of IBM server solutions.

The team that wrote this paper

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

Giuliano Anselmi has worked with IBM Power Systems for 15 years, and has a deep knowledge of the related hardware and solutions. He was previously a pSeries® Systems Product Engineer for 7 years, supporting the Web Server Sales Organization, IBM Sales, Business Partners, and Technical Support Organizations. He joined the Field Technical Sales Support group in 2004 and was accredited as an IT specialist in 2007. Giuliano is currently a system architect in IBM STG and supports the General Business division.

YoungHoon Cho is a System p Product Engineer at the Power Systems post-sales Technical Support Team in IBM Korea. He has seven years of experience working on RS/6000® and System p products. He is an IBM Certified Specialist in System p and AIX 5L™. He provides second line technical support to field engineers on System p and system management.

Jim Cook is an ITSO Rochester Project Leader and has worked with System i hardware and IBM i (i5/OS) configurations for over 20 years. Jim has worked on many ITSO redbooks and redpaper publications covering most areas of System i-based capabilities and has also been a project leader on the multiple sets of System i announcement presentations, known to System i IBMers and business partners as the "Technical Overview presentation set," available at the System i technical support Web site.

Gregor Linzmeier is an IBM Advisory IT Specialist for IBM System p workstation and entry servers as part of the Systems and Technology Group in Mainz, Germany. He supports 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, IBM IntelliStation® POWER®, and AIX in large CATIA client/server projects. He has special expertise in AIX Thin Server, Partition migration and Green IT.

Marcos Quezada is a Senior Accredited IT Specialist in Argentina. He has 10 years of IT experience as a UNIX® systems pre-sales specialist and as a Web Project Manager. He holds a degree on Informatics Engineering from Fundación Universidad de Belgrano. His areas of expertise include IBM RS/6000, IBM eServer™ pSeries/p5 and Power Systems servers under the AIX operating system and pre-sales support of IBM Software, SAP® and Oracle® solutions architecture running on IBM UNIX Systems with focus on competitive accounts.

John T Schmidt is an Accredited IT Specialist for IBM and has over 7 years experience with IBM and System p. He has a degree in Electrical Engineering from the University of Missouri - Rolla and an MBA from Washington University in St. Louis. He is currently working in the United States as a presales Field Technical Sales Specialist for System p in St. Louis, MO.

Guido Somers is a Cross Systems Certified IT Specialist working for IBM Belgium. He has 13 years of experience in the Information Technology field, ten years of which were within IBM. He holds degrees in Biotechnology, Business Administration, Chemistry, and Electronics, and did research in the field of Theoretical Physics. His areas of expertise include AIX, Linux, system performance and tuning, logical partitioning, virtualization, IBM PowerHA™, SAN, IBM System p servers, as well as other IBM hardware offerings. He currently works as a Client IT Architect for Infrastructure and Global ISV Solutions in the e-Business Solutions Technical Support (eTS) organization. He is also an author of the second edition of *Integrated Virtualization Manager on IBM System p5*, REDP-4061 and *PowerVM Live Partition Mobility on IBM System p*, SG24-7460.

Become a published author

Join us for a two- to six-week residency program! Help write a book dealing with specific products or solutions, while getting hands-on experience with leading-edge technologies. You will have the opportunity to team with IBM technical professionals, Business Partners, and Clients.

Your efforts will help increase product acceptance and client satisfaction. As a bonus, you will 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 paper or other IBM Redbooks® in one of the following ways:

► Use the online **Contact us** review Redbooks form found at:

 ibm.com/redbooks

► Send your comments in an e-mail to:

 redbooks@us.ibm.com

► Mail your comments to:

 IBM Corporation, International Technical Support Organization
 Dept. HYTD Mail Station P099
 2455 South Road
 Poughkeepsie, NY 12601-5400

General description

During 2008 IBM System i and IBM System p platforms unified the value of their servers into a single, powerful lineup of servers based on industry-leading POWER6 processor technology with support for the IBM AIX, IBM i (formerly known as i5/OS), and Linux operating systems. This new, single portfolio of IBM Power Systems servers offers industry-leading technology, continued IBM innovation, and the flexibility to deploy the operating system that your business requires.

The Power 570 and Power 595 models announced in April 2008 were announced as fully "converged" or "unified" at that time. The Power 520 and Power 550 models announced January 2008 and the April 2008 announcements brought these models very close to complete convergence. The October 2008 announcements, together with system firmware made available November 2008 brought full unification for the Power 520 and Power 550 systems.

This publication provides comprehensive overview-level technical information on the IBM Power 520 POWER6 and POWER6+ technology Power 520 Express™ servers with the following Machine Type and Model (MTM) values:

► 8203-E4A (1-core, 2-core, and 4-core configurations): Runs AIX, IBM i, and Linux operating systems. To run IBM i on the 8203-E4A you need system firmware level EL340_039 or later.

With the availability of the unified 8203-E4A or 8204-E8A configurations, you select the MTM first and then specify the primary operating system and any secondary operating systems. Specifying the primary operating system generates a $214n$ specify code:

2145   IBM i primary OS

2146   AIX primary OS

2147   Linux primary OS

The primary operating system $214n$ value is used as a Manufacturing Routing indicator and does not deliver parts, software, or services. It does, however, assist the IBM configurator in determining which features are shown by default on succeeding configuration windows. You can also indicate on the order any partition configurations you want IBM manufacturing to know about so that hardware feature cards are populated within the various processor and I/O enclosures that will be delivered to the customer ready to deploy the partition definitions.

This eliminates or significantly reduces the need to move adapter cards around to satisfy the partition definition.

The IBM System Planning Tool (SPT) can be used to pre-configure a valid system with or without partitions. A file representing the partition configuration can be exported and used as input to the IBM configurator.

For more information on SPT, refer to:

http://www.ibm.com/systems/support/tools/systemplanningtool

The System Planning Tool is discussed in 3.5, "System Planning Tool" on page 154.

> **Note:** The word *express* indicates a level of hardware and software pre-packaging (a starting configuration) to simplify ordering of hardware and in some cases a selected set of frequently used licensed programs with the intent of getting your operational environment up and running in a short time. In this publication we generally use the single term *Power 520,* dropping the word express when referring to the entire range of Power 520 hardware configuration options. This represents all Power 520 MTMs.
>
> These express edition configurations are slightly different, depending upon which operating system (AIX, IBM i, or Linux) is specified as the primary operating system.
>
> There are also solution editions offering certified applications from SAP and Oracle that are installed on a configured Power 520 ready to run. The hardware configuration of a Power 520 solution edition is based upon an express configuration.
>
> Once the order has been generated and the system delivered, the *express edition* or *solution edition* is not visible on the system.

The POWER6- and POWER6+-based Power 520 MTMs are designed to continue the tradition of the IBM POWER5™ and the IBM POWER5+™ processor-based System i 515, 520, and 525 models and the System p5 520 and 520Q Express servers.

The Power 520 model processor, memory, I/O loop attachment and many I/O device technologies are common across the operating systems supported.

Operating system-specific requirements and current System i and System p customer environments dictate some differences, which are documented where appropriate in this publication. Most support differences are in the area of I/O adapters and devices supported. We discuss I/O differences later in this chapter.

This chapter introduces many of the common architecture facilities and discusses some operating system support differences. Chapter 2, "Architecture and technical overview" on page 57 expands on this information and lists the I/O features supported by the three operating systems.

The IBM Power 520 servers all support 4.2 GHz POWER6 processors and 4.7 GHz POWER6+ processors.

All machine model types can be physically packaged as either a deskside or a 4U (Electronic Industries Alliance (EIA) Units) rack-mount server. In either configuration the Power 520 models deliver outstanding performance at low prices and can utilize one or a combination of the supported IBM AIX, IBM i, and supported Linux operating systems.

The Power 520 supports single- or dual-core POWER6 modules on a single- or dual-chip planar. Each POWER6 2-core module has 8 MB of L2 cache (4 MB per core). A 2- and 4-core

POWER6+ option is also available. System DDR2 SDRAM memory capacity starts at 1 GB up to 16 GB (1-core configuration), up to 32 GB (2-core configuration) and up to 64 GB (4-core configuration). 64 GB of memory is achieved using eight DDR2 memory DIMM slots.

> **Note:** Unless otherwise noted in the text, the term POWER6 in this publication includes both POWER6 and POWER6+ processor technologies.

The system unit has five Peripheral Component Interface (PCI) slots (three PCIe 8x and two PCI-X 64-bit (Double Data Rate) DDR 2.0) and six 3.5 inch internal SAS disk bays. The system unit must be ordered with one of the following Integrated Virtual Ethernet (IVE) options:

- ► Two Ethernet 10/100/1000 Mbps ports
- ► Four Ethernet 10/100/1000 Mbps
- ► Two 10 Gigabit Ethernet ports (not available on 9407-M15, 9408-M25, 9409-M50)

A slim media bay is available for a DVD-ROM or DVD-RAM, and a half high media bay is available for a tape drive.

Additional I/O adapters can be plugged into the system unit PCI-X DDR2 and PCIe card slots. I/O adapters and devices can also be added in up to either twelve or eight optional I/O enclosures (drawers or towers) using the up to two GX+ I/O loop adapters. IBM i supports up to six I/O enclosures and AIX and Linux support up to four, on a RIO-2 loop. AIX, IBM i, and Linux operating systems support up to four I/O enclosures on a 12x loop.

See Chapter 2, "Architecture and technical overview" on page 57, for more information on the IBM Power 520 POWER6 processor core, memory, and I/O attachment technology.

I/O Hardware Terminology Notes:

System i and System p customers have become familiar with terms that are different between the two customer sets. This needs to be addressed early in this publication to make it easier to understand the new hardware features as well as, in the case of I/O support, features currently available on POWER5 technology servers and supported on POWER6 technology-based servers.

► HSL and RIO terminology:

  System i has used HSL and System p has used RIO as different terms for the same I/O loop attachment technology. The POWER5 and POWER6 technology systems use the second generation of this technology and thus HSL-2 or RIO-2 terms are now used. In this publication we use RIO-2 in most cases. If you see HSL-2 it is the same loop technology as RIO-2.

  In some earlier publications System p documents may also use RIO-G instead of RIO-2.

► 12x terminology:

  12x is a newer and faster technology compared to RIO-2 high speed I/O loop technology. The *12* refers to the number of wires within the 12x cable. Potentially, 12x technology offers up to 50 percent more bandwidth than HSL technology. IBM's 12x loop technology is based upon the participation of IBM with the InfiniBand® Trade Association (IBTA). IBM 12x implementation is not 100 percent InfiniBand compliant. Therefore this publication does not use 12x and InfiniBand terms interchangeably.

  RIO-2 I/O enclosures must be attached to a RIO-2 loop. 12x I/O enclosures must be attached to a 12x loop. RIO-2 and 12x enclosures cannot be mixed on the same loop because they are not compatible. RIO-2 and 12x cables are different.

► PCIe terminology:

  PCIe uses a term called *lanes* to refer to its capabilities. Each lane can support a data rate of 2.5 Gbps for both send and receive. The slowest and smallest PCIe comes in one lane (referred to as X1) with the fastest PCIe up to 32 lanes (x32) and beyond. Think of lanes as data paths. The more lanes there are, the faster the data can flow, which is similar to an auto highway with multiple driving lanes. The size of the adapter and slot vary proportionally to the number of lanes. Since the physical size varies for both the adapter and slot, if an adapter has more lanes than a slot, the PCIe adapter cannot physically be plugged to that slot. In the reverse case, if the slot has more lanes than the adapter, then the adapter *can* be plugged into that slot.

  In summary, PCIe adapters are supported in a slot with an equal or higher number of lanes. When you see x8 associated with a PCIe card slot or adapter, that means the card slot or adapter supports eight lanes.

► I/O adapter or I/O controller terminology:

  You will see the word *adapter* and the word *controller* as the title of various I/O feature numbers, each of which supports attachment of other I/O hardware, such as tape or disk devices or a communications (LAN or WAN) cable. The words adapter and controller refer to the same general capability, and thus should be considered to be the same thing. In some System i I/O documentation, you may also see the acronym *IOA*. This is an acronym for an I/O adapter.

► System i documentation has a history of referring to a hardware component called an I/O Processor or IOP. These are used as a front-end interface to older technology IOAs, providing support or efficiencies not available in the older IOAs. See "AIX, IBM i, Linux I/O considerations" on page 15 for more information on IOPs.

As with POWER5-based technology, simultaneous multithreading (SMT) enabling two threads to be executed at the same time on a single physical processor core is a standard feature of POWER6 technology. SMT takes full advantage of POWER6 processor speeds and increased L2 and L3 (where available) cache sizes and interfaces.

Introduced with the POWER6 processor design is integrated hardware decimal floating-point support. On many industry-wide servers decimal floating point is implemented through software. Hardware support can improve the accuracy and performance of the basic mathematical calculations of financial transactions that occur regularly on today's business computers running C, C++, and Java™ applications. More information about this topic is in 2.1.1, "Decimal floating point" on page 60.

POWER6 also includes an AltiVec SIMD accelerator, which helps to improve the performance of high performance computing (HPC) workloads.

The IBM POWER6 architecture with EnergyScale technology provides features such as power trending, power saving, capping of maximum power, and thermal measurement. These features are enabled using IBM Systems Director Active Energy Manager™ software within the IBM Systems Director product's console component. With the Active Energy Manager, you can measure energy use and direct policies toward the energy-efficient operation of the server, while the underlying hardware automatically adjusts to deliver the desired operating solution. The POWER6 chip is also designed to conserve power and reduce heat generated by the server. A feature called *nap* mode enables processor clocks to be dynamically turned off when there is no workload to be done and turned back on when there are instructions to be executed.

See 2.2, "IBM EnergyScale technology" on page 62, for more information.

All Power Systems servers (POWER5 and POWER6 technology) include the POWER Hypervisor™, which provides the following basic virtualization capabilities:

► Logical partitioning (LPAR) technology.

► Dynamic LPAR: Dynamic LPAR (DLPAR) allows clients to dynamically allocate many system resources to application partitions without rebooting, with up to four dedicated processor partitions on a fully configured Power 520 system.

► VIrtual LAN: This is memory to memory inter-partition communication using a virtual Ethernet configuration.

In addition to the base virtualization that is standard on every Power 520 server, three optional virtualization features are available on the server:

► PowerVM Express Edition
► PowerVM Standard Edition (formerly Advanced POWER Virtualization (APV) Standard)
► PowerVM Enterprise Edition (formerly APV Enterprise)

These are managed using built-in Integrated Virtualization Manager (IVM) software or optionally through use of a hardware management console (HMC).

The PowerVM Express Edition and IVM allow users to create up to three partitions on the server, leverage virtualized disk and optical devices, utilize shared- or dedicated-capacity processor configuration, and use the Shared Processor Pool. A browser-based interface, IVM is used to manage virtual devices, processors and partitions. An HMC is not required. PowerVM Express Edition also includes PowerVM Lx86.

PowerVM Standard Edition includes all of the capabilities of PowerVM Express Edition plus capped or uncapped partition configuration, full function IBM Micro-Partitioning™, and Virtual I/O Server (VIOS) capabilities. For low processor utilization workloads, micro-partitions can

be defined as small as 1/10th of a processor. Processor capacity can be changed in increments as small as 1/100th of a processor. Up to 40 micro-partitions can be created on a 4-core system (10 per core). VIOS allows for the sharing of disk and optical devices and communications and Fibre Channel adapters. Also included is support for Multiple Shared Processor Pools and Shared or Dedicated processor Capacity partitions.

PowerVM Enterprise Edition includes all features of PowerVM Standard Edition plus Live Partition Mobility, available on POWER6 systems. Live Partition Mobility can be used by AIX or Linux partitions to allow a partition to be relocated from one server to another while end users are using applications running in the partition.

If installed, PowerVM licensing is required for all active processors.

For more information about PowerVM capabilities on the various POWER6 MTMs and the required operating system level supporting these capabilities, refer to Chapter 3, "Virtualization" on page 133.

AIX versions 5.3 and 6.1 support the POWER6 servers. Version 6.1 extends the capabilities of AIX to include new virtualization approaches, new security features, and new continuous availability features. It introduces a software-based virtualization approach called PowerVM AIX 6 Workload Partitions (WPARs) which enables the creation of multiple virtual AIX V6.1 environments inside of a single AIX V6.1 instance. Also available is a new licensed program product, the PowerVM AIX 6 Workload Partitions Manager™ for AIX (WPAR Manager). It allows management of WPARs across multiple systems, including the movement of PowerVM Workload Partitions from one system to another without restarting the application or causing significant disruption using PowerVM Live Application Mobility, a feature of AIX V6.1.

The IBM i operating system (formerly known as i5/OS) releases 5.4 and 6.1 run on POWER6 technology systems. If IBM i 5.4 is to be used, machine code level V5R4M5 is required. Both releases deliver a wide range of functions for customers familiar with the i5/OS operating system from a previous release.

For more specific information about the operating system levels that support these POWER6 server MTMs, see 1.16, "Operating system support" on page 51.

Additional features introduced with POWER6 processor-based technology include an Integrated Virtual Ethernet adapter standard with every system, the Processor Instruction Retry feature automatically monitoring the POWER6 processor and, if needed, restarting the processor workload without disruption to the application, and a new Hardware Management Console (HMC) graphical user interface offering enhanced systems control.

See Chapter 2, "Architecture and technical overview" on page 57 for more information about these topics.

1.1 System unit specifications

As discussed earlier, this publication uses the Power 520 MTM value of 8203-E4A to represent the unified Power 520 configuration that, starting in October 2008, can be ordered to run AIX, IBM i, or Linux operating systems. A system can be ordered to run all resources under only one operating system or configured to run multiple operating systems, each in its own partition.

The 8203-E4A can have a 1-core, 2-core, or 4-core configuration with all processors permanently activated.

The one-core 8203-E4A configuration has some I/O restrictions, such as not supporting any RIO-2 or 12X I/O loops.

Table 1-1 lists the general system specifications of the system unit. Note that in some publications the system unit might also be identified by the term processor enclosure or even the term CEC (Central Electronics Complex).

Table 1-1   System unit specifications

Description	Range (operating)
Operating temperature range	10° to 38° C (50 to 104 F)
Operating temperature recommended	20° to 28° C (68 to 82 F)
Relative humidity	8% to 80%
Maximum dew point (operating	17°C (62.6°F)
Noise level	Deskside system: 6.4 bels idle / 6.5 bels operating Rack-mount drawer: 6.8 bels idle / 6.8 bels operating
Operating voltage	100 to 127 or 200 to 240 V ac at 50/60 Hz (auto-ranging)
Maximum power consumption	950 watts (maximum)
Maximum power source loading	0.876 kVa (4-Way); 0.794 (2-Way)
Maximum thermal output	3242 BTU[1]/hr (maximum)
Maximum altitude	3,048 m (10,000 ft)

1 British Thermal Unit

1.2 Physical package

The system is available in both a rack-mounted and deskside form. The major physical attributes for each are discussed in the following sections.

Note, the expression #*nnnn* will be used to represent an orderable feature code in this publication.

1.2.1 Deskside model

Table 1-2 on page 8 provides a list of physical attributes for the deskside model and Figure 1-1 on page 8 shows a picture of the deskside model.

Table 1-2   Physical packaging of deskside model

Dimension	One CEC drawer
Height	540 mm (21.3 in)
Width	328.5 mm (12.9 in) with tip foot 182.3 mm (7.2 in) without tip foot
Depth	628 mm (24.7 in)
Weight	40.8 kg (90 lb) 37.6 kg *83 lb) without tip foot

.

Figure 1-1   Deskside model

1.2.2 Rack-mount model

The system can be configured as a 4U (4 EIA) IBM rack-mount by selecting:

▶ #7200 or an OEM rack-mount with #7201 for the 8203-E4A

Table 1-3 provides a list of physical attributes for the rack-mount model; Figure 1-2 on page 9 is a picture of the rack-mount model.

Table 1-3   Physical packaging of rack-mount drawer

Dimension	One CEC drawer
Height	173 mm (6.8 in)
Width	440 mm (17.3 in)
Depth	538 mm (21.2 in)
Weight	34.0 kg (75 lb)

Figure 1-2   Rack-mount model

The front and rear views of the system are shown in Figure 1-3.
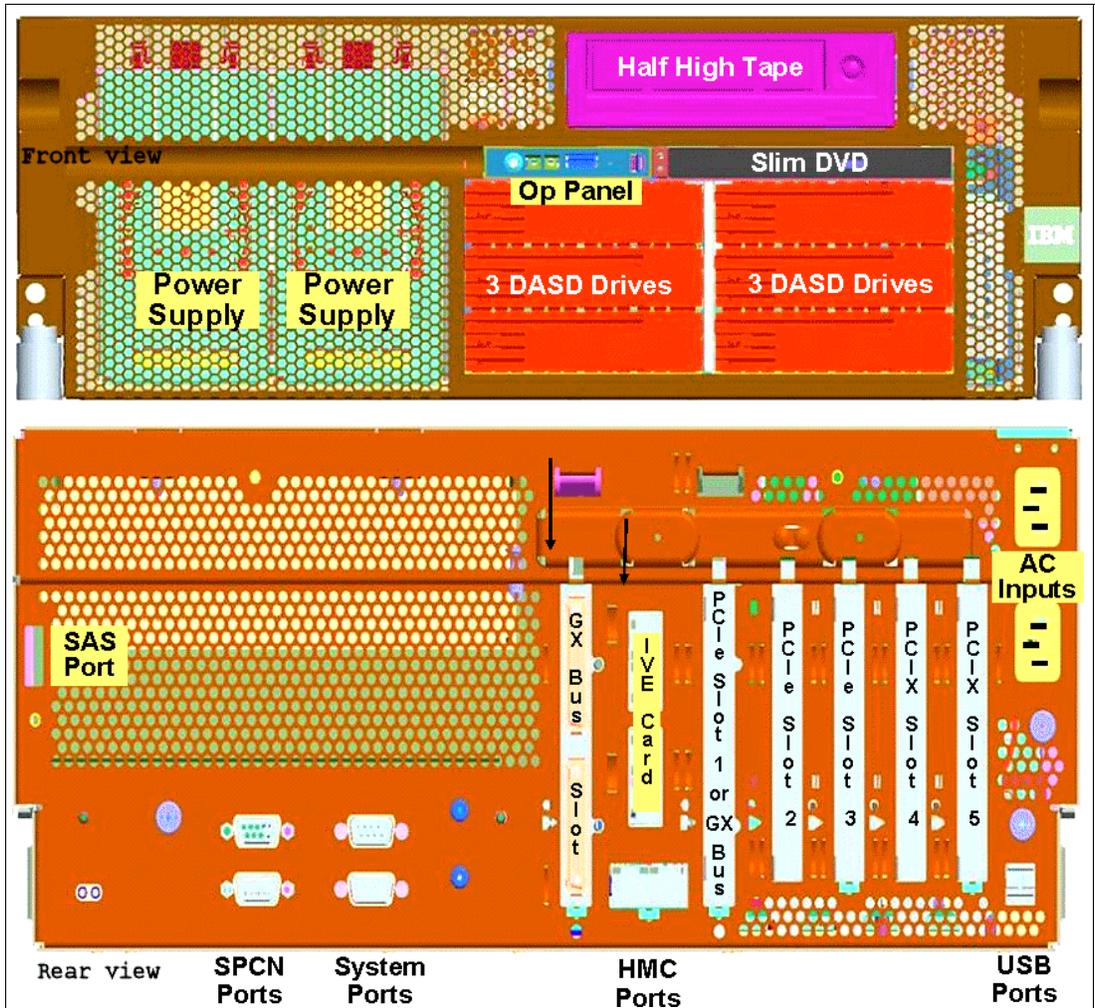
Figure 1-3   Front and rear system unit views

Figure 1-4 shows some additional details from a top rear view.

Power 520 -- Rear

POWER6 chip
• 1 or 2 cores

175MB write cache
• Optional
• For performance & RAID
• Includes aux write cache
• Has hot-plug battery

Fans
• Redundant
• Hot-swap

SAS port
• Attach EXP 12S Disk Drawer 2-core 520
• 175MB write cache pre-requisite

GX adapter Slot
• 2-core 520 only
• 12X or HSL/RIO loop
• Uses PCIe slot #1

Five PCI slots
• 2 PCI-X DDR
• 3 PCIe

Serial ports
• IBM i use for UPS comm connection

IVE Daughter Card
• 2 or 4 Ethernet 1Gb ports

AC Input

Figure 1-4   Rear top view

1.3 Minimum and optional system unit features

The system utilizes a 4.2 GHz POWER6 or 4.7 GHz POWER6+ processors available in 1-, 2-, and 4-core configurations. It features:

► 64-bit scalability, offering 1-, 2-, and 4-core configurations.

► 1 GB memory expandable to 64 GB.

– 16 GB maximum with a 1 processor configuration

– 32 GB maximum with a 2 processor configuration

– 64 GB maximum with a 4 processor configuration

► Up to 6 SAS DASD internal disk drives.

► Choice of DASD backplane options:

– 6 x 3.5 inch DASD backplane with no external SAS port (default)

– 6 x 3.5 inch DASD backplane with an external SAS port (optional)

– 8 x 2.5 inch DASD backplane with an external SAS port (optional)

► A maximum of five hot-swappable slots:

– Two PCIe x8 slots, short card length

– One PCIe x8 slot, full card length

– Two PCI X DDR slots, full card length

The slot positions are:

- Slot 1: PCIe x8 2.5 GHz short-length slot. A GX+ slot shares this slot.
- Slot 2: PCIe x8 2.5 GHz short-length slot
- Slot 3: PCIe x8 2.5 GHz full-length slot
- Slots 4 and 5: PCI X DDR 266 MHz full-length slots.

► Up to 2 GX+ loop adapter slots:

- One GX+ slot, shares PCIe x8 slot 1 (not available on a 1-core configuration)
- One GX+ slot, unshared (available only on a 4-core configuration)

These optional GX+ adapters (RIO-2 and, or 12x) are used for external to the system unit disks and I/O enclosure expansion.

► Two media bays:

- One slim bay for a DVD-ROM (optional) or DVD-RAM (optional)
- One half-high bay for a tape drive (optional)

Each system includes the following native ports:

► Choice of Integrated Virtual Ethernet I/O options:

- 2-port 1 Gigabit Integrated Virtual Ethernet (single controller)
- 4-port 1 Gigabit Integrated Virtual Ethernet (single controller)
- 2-port 10 Gigabit Integrated Virtual Ethernet (single controller). As stated earlier the 10 Gbps adapter is not supported on 9407-M15 or a 9408-M25.

► Three USB ports.

► Two system (serial) ports. One of these parts can be used by IBM i for Uninterruptible Power Supply (UPS) communication cable connection.

► Two HMC ports.

► Two SPCN ports.

► One Power Supply, 950 Watt AC, Hot-swappable, Base (redundant power optional).

► Redundant hot-swappable cooling fans.

1.3.1 Processor card features

The server is available in one, two, or four 4.2 GHz POWER6 or two and four 4.7 GHz POWER6+ processor card configurations, but you can use Capacity on Demand to activate an unactivated processor core. Because the processors are soldered onto the planar, there is no processor upgrade capability. Each processor card contains 64 KB I-cache (instruction cache), 64 KB D-cache (data cache), and 4 MB of Level 2 cache per core.

The Power 520 has permanent processor on demand capability. Other capacity on demand capabilities, such as (temporary) On/Off, Reserve, Trial, or Utility CoD features are available on POWER6 570 and 595 servers.

Table 1-4 on page 12 identifies the feature codes for processor cards at the time of writing.

Table 1-4   Processor card feature codes

Processor card feature	Description
5633 ► 5651 ► 5676	1-core 4.2 GHz POWER6 Processor Card, 4 Memory DIMM Slots ► One Processor Activation for Processor #5633 ► Zero-priced Processor Activation of #5633
5634 ► 5654 ► 5677	2-core 4.2 GHz POWER6 Processor Card, 4 Memory DIMM Slots ► One Processor Activation for Processor #5634 ► Zero-priced Processor Activation for #5634
5635 ► 5655 ► 5678	4-core 4.2 GHz POWER6 Processor Card, 4 Memory DIMM Slots ► One Processor Activation for Processor #5635 ► Zero-priced Processor Activation #5635
5577 ► 5578 ► 5579	2-core 4.7 GHz POWER6+ Processor Card, 4 Memory DIMM Slots ► One Processor Activation for Processor #5577 ► Zero-priced Processor Activation for #5577
5587 ► 5588 ► 5589	4-core 4.7 GHz POWER6+ Processor Card, 8 Memory DIMM Slots ► One Processor Activation for Processor #5587 ► Zero-priced Processor Activation #5587

1.3.2 Memory features

The amount and type of memory supported on the systems depends on the processor core configuration installed. A minimum 1 GB of memory is required. Four memory DIMM slots are available on a 1-core configuration, and the maximum system memory is 16 GB. The system memory feature codes cannot be mixed. Feature #4524 (16384 MB memory feature) is not available on a 1-core configuration.

Four memory DIMM slots are available on a 2-core configuration. The maximum system memory is 32 GB. The system memory feature codes cannot be mixed.

Memory feature #4288 is not available on a 1-core or 2-core configuration.

Eight memory DIMM slots are available on a 4-core configuration. The maximum system memory is 64 GB. There are two system locations where system memory can be installed, four DIMM slots in each location.

System memory must be of the same speed and size within the same processor card. Memory on a second card may be a different size and feature, but it still must be within the same speed and size as before. Memory installed evenly across all processor cards in the system typically results in the best possible performance for that configuration.

Plans for future memory upgrades should be taken into account when deciding which memory feature size to use at the time of initial system order. For example, if you initially order smaller memory DIMMs and later want to add memory, you might find that to achieve your desired maximum memory requires discarding the originally ordered DIMMs.

Table 1-5 shows the memory feature codes that are available at the time of writing for the 8203-E4A server. A 4-core configuration 8203-E4A is required to support the largest memory configuration.

Table 1-5   Memory feature codes

Feature code	Description	Model supported
4288	Memory Offering, 64 GB (Multiples of 4 of 2x8 GB DIMMs)	4-core models

Feature code	Description	Model supported
4520	1024 MB (2x512 MB) RDIMMs, 667 MHz, 512 MB DRAM	1-, 2- or 4-core models
4521	2048 MB (2x1024 MB) RDIMMs, 667 MHz, 512 MB DRAM	1-, 2- or 4-core models
4522	4096 MB (2x2048 MB) RDIMMs, 667 MHz, 512 MB DRAM	1-, 2- or 4-core models
4523	8192 MB (2x4096 MB) RDIMMs, 667 MHz, 512 MB Stacked DRAM	2- or 4-core models
4524	16384 MB (2x8192 MB) RDIMMs, 400 MHz, 1 GB Stacked DRAM	2- or 4-core models
4532	4 GB (2x2GB) DIMMs, 667MHz, 1 GB DRAM	1-, 2-, or 4-core models

> **Note:** If you mix memory cards of varying speeds (that is, different in location 1 than location 2) all memory functions at the lowest speed.

1.3.3 System unit disk and media features

This topic focuses on the I/O device support within the system unit. Each system unit features one SAS DASD controller with six hot-swappable 3.5 inch SAS disk bays and one hot-plug, slim-line media bay per enclosure. Only the new SAS DASD disk drives are supported within the system unit. There are two DASD backplane options, listed in Table 1-6.

SCSI technology disk drives are supported, but must be located in an I/O enclosure attached to a GX adapter loop.

Table 1-6   SAS DASD backplane options

Feature code	Description
8308	6 x 3.5 inch DASD backplane with no external SAS port (default)
8310	6 x 3.5 inch DASD backplane with an external SAS port (optional)
8346	8 x 2.5 inch DASD backplane with an external SAS port (optional),

Note that feature #8308 is the default, base function backplane with no external SAS port. It provides a single path from the SAS controller to each internal disk drive. #8310 an #8346 are the optional, higher function DASD backplanes with an external SAS port. They provide dual paths from the SAS controller to each internal disk drive using SAS expanders. This option, along with #5679, provides for internal RAID levels 0, 5, 6, and 10.

Features #8310 and #8346 also provide an external SAS connector for attachment of a single #5886, EXP 12S SAS Disk Drawer. This is an expansion drawer with 12 SAS storage slots. Feature #5886 supports up to 12 hot-swap SAS disk drives in mega-pack carriers. The EXP 12S is described in 2.12, "5886 EXP 12S SAS DASD Expansion Drawer" on page 97.

Selection of #8310 and #8346 requires SAS Cable, DASD Backplane to Rear Bulkhead, #3668.

> **Note:** To take advantage of advanced features, whether for functions with current or future availability, #8310 or #8346 should be selected to avoid the need to replace the #8308 backplane at a later time.

Figure 1-5 on page 14 shows an example of a Power 520 configuration that includes the EXP 12S SAS Disk Drawer attached to the system unit using #5679 and the #8310 backplane.

Sample configuration of internal disk drives

#5886 EXP 12S
Disk Drawer

POWER6 520

8310 Backplane

SAS cable

#5679 175MB write cache
for imbedded SAS controller
with auxiliary write cache

2U space

RAID-5/6
Or Mirror

4U space

RAID-5/6
Or Mirror

- *Up to 18 SAS disks*
  - 6 plus 12
- *One controller with*
  175MB write cache
- *Protected write cache*
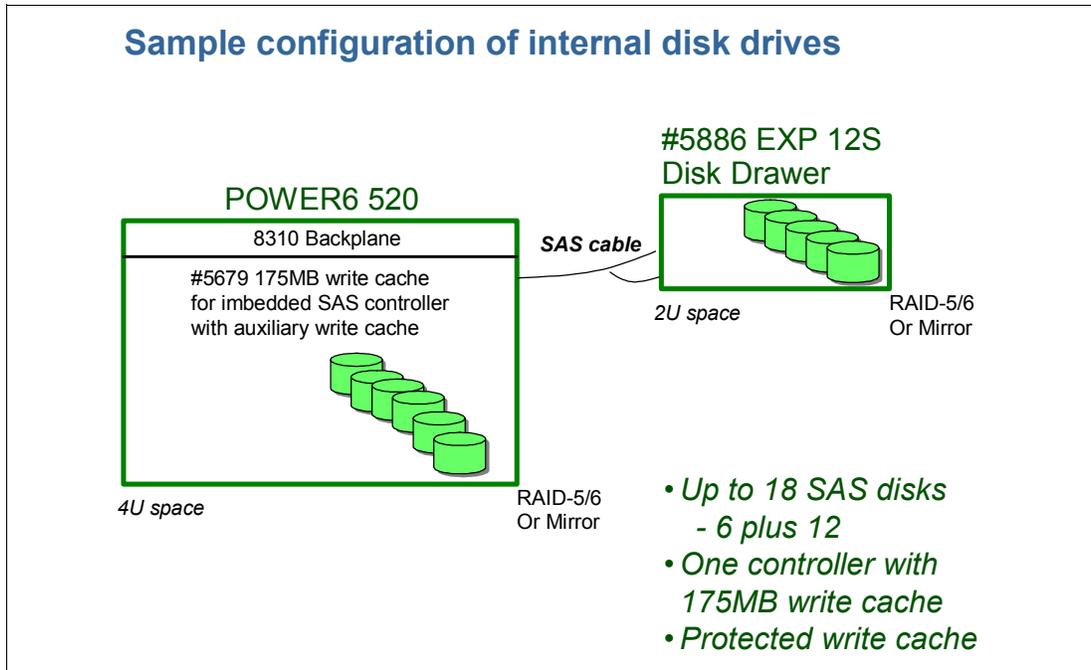
Figure 1-5   EXP 12S attached to the system unit example

The #5679 does not consume a PCI slot. It is two cards. The first card provides RAID support
and a 175 MB write cache; the second card is the auxiliary 175 MB write cache and battery.
With no performance impact, the auxiliary write cache contains a copy of the first card's write
cache data at all times. If the first card's write cache requires service, the auxiliary write cache
contents can be restored (by IBM service personnel) to a replacement write cache on the first
card.

Notes:

► The cache backup capabilities are supported regardless of whether RAID or IBM i
operating system mirroring is in effect.

► The write cache battery is hot-pluggable.

► The IBM i operating system supports other disk adapter features that include large write
cache with auxiliary write cache. These adapters are supported in either the system
unit PCI slots or PCI slots within the I/O enclosures attached using a GX+ adapter loop.
These adapters are described in more detail in 2.8.3, "SCSI and SAS adapters" on
page 79.

Currently there are several physical capacity SAS disk drives that can be placed within the
system unit or the ESP™ 12S. Depending on whether the disk is ordered new and formatted
for AIX or Linux, or the IBM i operating system models, there are different orderable feature
numbers. There are also different maximum data capacities available because the IBM i
operating system uses eight additional bytes for virtual storage integrity information per disk
sector.

Table 1-7 on page 15 lists the disk drive feature codes that each bay can contain. Note that
disks formatted for AIX or Linux cannot be used directly by IBM i. IBM i 6.1 can be a client
partition to a VIOS partition that accesses the AIX or Linux formatted disks. Disks formatted
for IBM i cannot be directly accessed by AIX or Linux. AIX or Linux can be a client partition to
IBM i 5.4 or 6.1, and can thus provide virtual disk support using the IBM i formatted disks.

Table 1-7   SAS disk drive feature code description: 8203-E4A and AIX, IBM i disks

Feature Code	Description	OS support: AIX, Linux, IBM i
1882	146.8 GB 15 K RPM SAS Disk Drive SFF 2.5 inch	A, L
1883	73.4 GB 15 K RPM SAS Disk Drive SFF 2.5 inch	A, L
1884	69.7 GB 15 K RPM SAS Disk Drive (2.5 inch)	i
1890	69 GB Solid® State Drive SFF (2.5 inch)	A, L
1909	69 GB Solid State Drive SFF (2.5 inch)	i
3586	69 GB Solid State Drive 3.5 inch	A, L
3587	69 GB Solid State Drive 3.5 inch	i
3647	146 GB 15 K RPM SAS Disk Drive	A, L
3648[2]	300 GB 15 K RPM SAS Disk Drive	A, L
3649	450 GB 15K RPM SAS Disk Drive	A, L
3658[1]	428 GB 15 K RPM SAS Disk Drive	i
3677	139.5 GB 15 K RPM SAS Disk Drive	i
3678[1,2]	283.7 GB 15 K RPM SAS Disk Drive	i
General note: A-AIX; L-Linux; i-IBM i 1. If used as IBM i load source device, you must use IBM i 6.1. 2. IBM has announced plans to withdraw these features from marketing during February 2009. IBM plans are subject to change without notice.		

The system has a slim media bay that can contain an optional DVD-ROM or an optional DVD-RAM, and a half-high bay that can contain a tape drive. Table 1-8 lists the media feature codes available. If a tape device is installed in the half-high media bay, #3655, SAS HH Cable must be selected.

Table 1-8   Media bay features

Feature code	Description	OS support: AIX, Linux, IBM i
5743	SATA Slimline DVD-ROM Drive	A, L, i
5762	SATA Slimline DVD-RAM Drive	A, L, i
5746	800 GB/1.6 TB Half High LTO4 SAS Tape Drive	A, L, i
5907	36/72 GB 4 mm DAT72 SAS Tape Drive	A, L, i
5689	DAT 160 Tape Cartridge 80/160 GB	A, L, i
5619	80/160 GB 4 mm SAS Tape Drive	A, L, i
5747	800 GB LTO-4 Tape Cartridge	A, L, i

1.3.4 AIX, IBM i, Linux I/O considerations

Previously we stated that there are some operating system-specific requirements and that current System i and System p customer environments dictate some differences in orderable configurations, which are documented where appropriate in this publication.

Examples of AIX unique I/O features include graphic adapters, specific WAN/LAN adapters, SAS disk/tape controllers, iSCSI adapters, and specific Fibre Channel adapters.

Examples of IBM i unique I/O features include the #5094/5294/5088/0588/0595 I/O drawers/towers (generically, *I/O enclosures*), I/O processors (IOPs), IOP-based PCI adapter cards, very large write cache disk controllers (adapters), specific Fibre Channel adapters, iSCSI adapters, and specific WAN/LAN adapters.

System i hardware technologies and the IBM i operating system (OS/400®, i5/OS, and so forth) have a long history of supporting I/O adapters (also commonly referred to as *controllers,* or simply *IOAs*) that also required a controlling I/O Processor (IOP) card. A single IOP might support multiple IOAs. The IOP card originally had a faster processor technology than its attached IOAs. Thus, microcode was placed in the IOP to deliver the fastest possible performance expected by customers.

There are two important examples of the need for an IOP in System i environments: SNA and tape support.

► SNA support:

There are a number of customers using SNA communications protocols, typically using the underlying Synchronous Data Link Control (SDLC) protocol. The IBM i implementation is to have IBM Systems Network Architecture (SNA) processes (and on remote WAN lines, SDLC protocol) within the IOP. The IOAs were not powerful enough to provide this and work in the IOP off-loaded work that would consume CPU cycles building up data transmission blocks and decoding incoming transmission blocks.

Newer IOAs have sufficient capacities, but with the SNA communications world no longer expanding, re-implementing SAN processing for the newer IOAs is not cost justified.

However, where possible, customers with SNA-based networks will generally want to keep the advantages available through the IOP-IOA relationship.

Alternatively, you can use the newer IOA communication adapters that do not work with an IOP and use operating system-provided SNA Enterprise Extender support. However, in that environment, you must ensure both ends of the conversation support SNA Enterprise Extender protocols.

Enterprise Extender frames and protocols are *masked* within IP frames so the data is freely routed across existing IP-based network. A lengthy discussion about Enterprise Extender support is beyond the scope of this publication. It does provide some advantages to SNA-based environments, but the key to its use is that both ends of the conversation must support Enterprise Extender capabilities.

► Tape support:

Some of the older tape technologies are supported by IBM i only when the tape device is attached to an IOA with a supporting IOP.

For additional information on IOAs and controlling IOPs that enable IBM i to support "SNA direct" (not using Enterpriese Extender support) and specific tape device or tape library support, refer to *IBM Power 520 and Power 550 (POWER6) System Builder*, REDP-4412. For tape support see the "Tape and optical storage attachment" chapter in REDP-4412.

Because the POWER6 system units do not support IOPs, the IOP-IOA configuration must be preserved within a supporting I/O enclosure attached to the system using a RIO-2 loop adapter.

IOAs introduced over the last two to three years have very fast processors and do not require a supporting IOP. Within the System i community these adapters are sometimes referred to

as *smart IOAs* or *dual mode IOAs* because they can operate with or without an IOP. There are also IOAs that do not run with an IOP; these are sometimes referred to as an *IOP-less IOAs*.

AIX or Linux client partitions hosted by an IBM i partition are not aware of any unique IBM i I/O hardware requirements.

For new system orders IOP-less IOAs are what AIX or Linux users consider as the standard I/O environment. New orders for AIX, IBM i, and Linux operating systems should specify the smart or IOP-less IOAs.

As discussed previously, using SNA and older tape device support, System i customers who move to the POWER6 models need to determine how to handle any existing IOP-IOA configurations they may have. Older technology IOAs and IOPs must be replaced or I/O enclosures supporting IOPs must be used.

Later in this publication we discuss the PCI technologies that can be placed within the processor enclosure. For complete PCI card placement guidance in a POWER6 configuration, including the system unit and I/O enclosures attached to loops, we refer you to the following documents available at the IBM Systems Hardware Information Center under the Power Systems category at:

http://publib.boulder.ibm.com/infocenter/systems/scope/hw/index.jsp

► For PCI placement information associated with the April 2008 POWER6-based IBM Power 520 9407-M15, 9408-M25, and 9409-M50 model configurations, refer to *Power Systems PCI Adapter Placement Guide for Machine Type 940x,* SA76-0096.

► For PCI placement information associated with the April 2008 POWER6-based IBM Power 570 and 595 model configurations, refer to *Power Systems PCI Adapter Placement Guide for Machine Type 820x and 91xx,* SA76-0090.

In recent years System i customers have used a different PCI placement publication, commonly referred to as the System i PCI card placement rules publication. The most recent PCI card placement System i publication is *PCI, PCI-X, PCI-X DDR, and PCIe Placement Rules for IBM System i Models*, REDP-4011. Though updated April 2008, this publication does not cover the 2008 POWER6-based I/O card placement details. There are no plans to update this System i publication. For POWER6 placement information you should use SA76-0096 or SA76-0090.

1.3.5 I/O enclosures attachment using 12x or RIO-2 I/O loop adapters

As discussed at the beginning of 1.3, "Minimum and optional system unit features" on page 10, the system unit has five PCI-based I/O expansion slots. There are two PCIe 8x short-length slots, one PCIe 8x full-length slot, and two PCI-X DDR long slots. The PCIe short slot 1 is shared with the first GX+ slot. A second GX+ slot is available on a 4-core configuration.

Chapter 2, "Architecture and technical overview" on page 57, discusses the hardware features that can be plugged into the PCIe and PCI-X DDR slots, as well as features that can be plugged into I/O enclosures supported on either a RIO-2 or 12x I/O loop through an adapter plugged into GX+ slot.

If more PCI slots are needed than can be contained by the system unit, optional GX adapters (not supported on 1-core configurations) can be installed, allowing up to two I/O loops with up to four or six I/O enclosures per loop to be ordered and configured. By the term *I/O*

enclosures we include the following I/O hardware containers associated with a specific orderable feature:

► I/O drawers
► I/O expansion units
► I/O towers

Existing System i and System p model configurations have a set of I/O enclosures that have been supported on RIO-2 (HSL-2) loops for a number of years. Most continue to be supported on POWER6 models. This section highlights the newer I/O enclosures that are supported by the POWER6 models that are actively marketed on new orders. Chapter 2, "Architecture and technical overview" on page 57 provides more information on all supported I/O hardware.

Table 1-9 lists the available GX+ adapter types and their feature numbers.

Table 1-9   GX adapters

Feature code	Description	OS support: AIX, Linux, IBM i
5609	GX Dual-port 12x (DDR) channel attach adapter	A, L, i
5614	Dual port RIO-2 I/O hub	A, L, i
5616	GX dual-port 12x (SDR) channel attach adapter	A, L, i

These GX adapters are not supported on the 9407-M15 or the 1-core 8203-E4A.

The IBM i and AIX operating systems each support a defined set of I/O enclosures and I/O adapters on the loop connected to the GX adapter. The enclosures supported are unique based upon RIO-2 or 12x loop technology and the operating system.

Each I/O enclosure supports a specific GX+ adapter technology listed in Table 1-9.

1.4 External disk subsystems

The IBM Power 550 (MTM 8204-E8A) has internal hot-swappable drives. When the AIX operating system is installed in a 8204-E8A server, the internal disks are usually used for the AIX rootvg volume group and paging space. Specific client requirements can be satisfied with the several external disk possibilities that the 8204-E8A supports.

For IBM i support of external disk subsystems, see Figure 1-6 on page 19.

	IBM i – directly supported		IBM i – indirectly supported via VIOS	
	POWER5	POWER6	POWER5	POWER6
DS8000	Y [1]	Y	N	Y
DS6000	Y [1]	Y [2]	N	N
DS5000	N	N	N	Y
DS4000	N	N	N	Y
DS3000	N	N	N	Y
N Series	Y (NFS only)	Y (NFS only)	N	N
SVC	N	N	N	Y
XIV	N	N	N	N

1. Without IOP for IBM 1 6.1 excluding load source
2. Without IOP for IBM 6.1

Figure 1-6   IBM i External Storage compatibility

The following sections address AIX and Linux support in this area.

1.4.1 IBM System Storage N3000, N5000, N6000, N7000, and N Series Gateway

The IBM System Storage™ N3000, N5000, N6000, and N7000 line of iSCSI-enabled storage offerings provide a flexible way to implement a Storage Area Network over an Ethernet network. Flexible Fibre Channel and SATA disk drive capabilities allow for deployment in multiple solution environments, including data retention, nearline storage, disk-to-disk backup scenarios, and high performance mission-critical I/O intensive operations.

The newest member of the IBM System Storage N series family is the N Series Gateway. The IBM System Storage N series Gateway product line is a network-based unified storage solution designed to provide Internet Protocol (IP) and Fibre Channel (FC) protocol access to SAN-attached heterogeneous storage arrays. The N6000 and N7000 series ordered with a Gateway feature code can help you make the most of the dynamic provisioning capabilities of Data ONTAP® software across your existing Fibre Channel SAN infrastructure to support an expanded set of business applications.

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

1.4.2 IBM TotalStorage Storage DS3000 Series

The DS3000 product line is the entry level of the DS storage family. Designed to deliver advanced functionality at a breakthrough price, these systems provide an exceptional solution for workgroup storage applications such as e-mail, file, print and Web servers, as well as collaborative databases and remote boot for diskless servers.

The IBM TotalStorage® DS3000 Storage server family consists of the following models:

► DS3200
► DS3300
► DS3400

The IBM System Storage DS3200 Express is an external storage enclosure specifically designed for the SMB. The DS3200 addresses the top concerns of these businesses managing increasing demand for capacity, data availability, and functionality. DS3200 scales up to 21.6 TB (using 3 EXP3000s and 48 450 GB SAS disks) and 48.0 TB (using 3 EXP3000s and 48 1.0 TB SATA disks).

The IBM System Storage DS3300 Express is an external storage system specifically designed for a range of organizations' data requirements. With iSCSI protocol, next-generation SAS back-end technology, and SAS and SATA drive intermix support, the DS3300 Express storage system provides businesses with robust, reliable, and cost-effective networked storage. It is expandable by attaching up to three EXP3000s for a total of 21.6 TB of storage capacity with 450 GB SAS or up to 48.0 TB with 1.0 TB SATA.

The IBM System Storage DS3400 Express is an FC host-based external storage system specifically designed for a wide range of organizations. With FC interface technology, next-generation SAS back-end technology, SAS and SATA drive intermix support, and DS3000 Storage Manager software, the DS3400 storage system provides businesses with robust, reliable, cost-effective FC networked storage. The enclosure addresses the top concerns of businesses managing increasing demand for capacity, data availability, and consolidated management. The DS3400 is expandable by attaching up to three EXP3000s for a total of 21.6 TB of storage capacity with 450 GB SAS or up to 48.0 TB with 1.0 TB SATA.

IBM i attachment is supported with the DS3400 only via IBM Virtual I/O Server (VIOS).

For support of additional features and for further information about the IBM TotalStorage DS3000 Storage Server family, refer to the following Web site:

http://www.ibm.com/servers/storage/disk/ds3000/index.html

1.4.3 IBM TotalStorage Storage DS4000 Series

IBM System Storage and System Storage DS4000® series are disk storage products using redundant array of independent disks (RAID) that contain the Fibre Channel (FC) interface to connect both the host systems and the disk drive enclosures. The DS4000 series of disk storage systems is an IBM solution for mid-range/departmental storage requirements. The products in the IBM System Storage DS4000 series family are described in the following paragraphs.

The DS4700 Express offers high performance 4 Gbps-capable Fibre Channel connections, up to 33.6 TB of Fibre Channel physical storage capacity, 112 TB of SATA physical storage capacity, and powerful system management, data management, and data protection features. The DS4700 Express is designed to expand from workgroup to enterprise-wide capability with up to six Fibre Channel expansion units with the DS4000 EXP810 Expansion Unit.

The IBM System Storage DS4800 disk storage system supports a high-performance 4 Gbps Fibre Channel interface. Increased host connectivity delivers the necessary bandwidth for high throughput applications. Designed for data-intensive applications that demand increased connectivity, eight 4 Gbps host channels can help provide up to 1724 MBps of sustained bandwidth, allowing for high throughput applications through eight channels directly attached to the host servers or connected to a Fibre Channel storage area network (SAN). The DS4800 can support up to 224 FC drives using EXP810, EXP710 or EXP100 Expansion Units. Four models are available: the 80A and 82A with 4 GB of cache, the 84A with 8 GB of cache, and the 88A with 16 GB of cache. All models support over 67.2 TB of FC physical storage capacity and 168 TB of Serial ATA (SATA).

IBM i attachment is supported with the DS4700 and DS4800 only via IBM Virtual I/O Server (VIOS).

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

1.4.4 IBM TotalStorage Storage DS5000 Series

The IBM System Storage DS5000 series storage system is designed to meet today's and tomorrow's demanding open-systems requirements while establishing a new standard for life cycle longevity. With its relentless performance and superior reliability and availability, the DS5000 series storage system can support the most demanding service level agreements (SLAs). And when requirements change, the DS5000 series can add or replace host interfaces, grow capacity, add cache, and be reconfigured on the fly—ensuring that it will keep pace with your growing company.

The IBM TotalStorage DS5000 Storage server family consists of the following models:
- ▶ DS5100
- ▶ DS5300

The DS5100 is the less featured of the two models and is targeted at high-end DS4000 customers. It has sixteen 4 Gbps FC drive interfaces and can hold a maximum of sixteen EXP5000 expansion units or a mix of DS5000 and EXP810 units for migration purposes, for a total of up to 256 disk drives. The DS5100 has a total of four 4 Gbps FC host ports on each of its two controllers and 8 GB of cache memory.

The DS5300 server has greater scalability than the DS5100. It has sixteen 4 Gbps FC drive interfaces and can hold a maximum of sixteen EXP5000 expansion units or a mix of EXP5000 and EXP810 units for migration purposes, for a total of up to 256 disk drives. It is designed to deliver data throughput of up to 400 MBps per drive port. The DS5300 has either four or eight 4 Gbps FC host ports on each controller and 8 or 16 GB of cache memory.

For support of additional features and for further information about the IBM TotalStorage DS5000 Storage Server family, refer to the following Web site:

http://www.ibm.com/servers/storage/disk/ds5000/index.html

1.4.5 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. DS6800 is designed to provide medium and large businesses with a low-cost, enterprise-class storage solution to help simplify data management and to provide comprehensive data protection and recovery capabilities and easy scalability for both mainframe and open system storage needs. It scales to 67.2 TB of physical storage capacity by adding storage expansion enclosures. The 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 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.

IBM i attachment is supported with the DS6800 and DS8000 directly and via IBM Virtual I/O Server (VIOS).

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

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

1.4.6 IBM System Storage SAN Volume Controller (SVC)

The IBM System Storage SAN Volume Controller (SVC) is a storage virtualization system that enables a single point of control for storage resources to help support improved business application availability and greater resource utilization. The objective is to manage storage resources in your IT infrastructure and to make sure they are used to the advantage of your business—and do it quickly, efficiently, in real time, while avoiding administrative cost.

SAN Volume Controller combines hardware and software into an integrated, modular solution that is highly scalable. An entry level SAN Volume Controller configuration contains a single I/O Group, can scale out to support four I/O Groups, and can scale up to support 1024 host servers and up to 8192 virtual disks. This configuration flexibility means that SAN Volume Controller configurations can start small with an attractive price to suit smaller environments or pilot projects and then can grow with your business to manage very large storage environments (up to eight petabytes).

IBM i attachment is supported with the SVC via IBM Virtual I/O Server (VIOS) only.

http://www-03.ibm.com/systems/storage/software/virtualization/svc/

1.4.7 IBM XIV Storage System

The IBM XIV® Storage System is a next-generation high-end open disk storage system. The XIV Storage System is an innovative, enterprise disk system based on a grid of standard, off-the-shelf hardware components. The IBM XIV system employs a storage architecture designed to provide the highest levels of performance, reliability, and functionality, combined with unprecedented ease of management and exceptionally low TCO.

The following list is a summary of important features offered by the IBM XIV Storage System:

► A revolutionary high end disk storage architecture designed to eliminate the complexity of administration and management of tiered storage and information life cycle management.

► Near-instantaneous and highly space-efficient snapshots provide point-in-time copies of data, which consume storage capacity only per changes while maintaining high performance.

► System virtualization that greatly simplifies IT operations and optimizes performance through automatic distribution of data across system resources, avoiding hot spots without manual tuning.

► High reliability achieved through unique self-healing functionality, which can enable the system to rebuild a 1 TB disk drive within 30 minutes or less, with almost no performance impact.

► Optimized, consistent performance derived from the system's massive parallelism, disk utilization, and unique caching algorithms.

► Intuitive user interface and system virtualization greatly simplify storage configuration and management.

► Built-in thin provisioning that can help reduce direct and indirect costs by allowing users to install capacity only for data actually written, and gradually grow it over time with minimal management effort.

► Greener power usage enabled through the use of large capacity SATA drives and optimized use of disk capacity, resulting in outstanding power consumption efficiency per TB, without compromising performance.

Customer-centric, low point of entry (27 TB usable) with incremental scaling (in 6 TB increments) to full rack capacity (79 TB usable), enabling organizations to start small based on current needs and flexibly add capacity while in production and with no need to reconfigure.

1.5 Hardware Management Console models

A Hardware Management Console (HMC) is supported but not required in some configurations of the Power 520 system. It is recommended to provide a set of functions to manage the system, including Logical Partitioning, inventory and microcode management, and remote power control functions. This is the default configuration for servers supporting logical partitions with dedicated or virtual I/O. Connection of an HMC disables the two integrated system ports.

For non-HMC systems, there are the following two service strategies:

► Full system partition: A single partition owns all the server resources and only one operating system can be installed. A console option other than the HMC must be specified on new orders. 8203-E4A, 9407-M15, 9408-M25 have different console options, depending upon whether AIX or IBM i is specified as the primary operating system and customer choice.

► Partitioned system: In this configuration, the system can have more than one partition and can be running more than one operating system. In this environment, partitions are managed by the Integrated Virtualization Manager (IVM), which provides some of the functions provided by the HMC. A console must be specified on new orders.

Table 1-10 lists the HMC models available for POWER6-based systems at the time of writing. They are preloaded with the required Licensed Machine Code Version 7 (#0962) to support POWER6 systems, in addition to POWER5 and POWER5+ systems.

Existing HMC models 7310 can be upgraded to Licensed Machine Code Version 7 to support environments that can include POWER5, POWER5+, and POWER6 processor-based servers. Version 7 is not available for the 7315 HMCs. Licensed Machine Code Version 6 (#0961) is not available for 7042 HMCs, and Licensed Machine Code Version 7 (#0962) is not available on new 7310 HMC orders.

Table 1-10   POWER6 HMC models available

Type-model	Description
7042-C07	IBM 7042 Model C07 desktop Hardware Management Console
7042-CR4	IBM 7042 Model CR4 rack-mount Hardware Management Console

Note: HMC level 7.3 or later is required. 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 you should never connect an HMC to a working IVM system. If you want to migrate an environment from IVM to HMC, you have to rebuild the configuration setup manually.

1.6 Build-to-order

You can perform a *build-to-order* or *a la carte* configuration using the IBM configurator where you specify each configuration feature that you want on the system. You build on top of the base required features, such as the embedded Integrated Virtual Ethernet 2-port or 4-port adapter.

IBM recommends that you start with one of several available "starting configurations," such as the Express Product Offering or Solution Editions. These solutions are available at initial system order time with a starting configuration that is ready to run as is.

1.7 IBM Power Systems Express Editions for AIX / Linux

The Express Product Offering is available only as an initial order. Express ordering enables easier initial ordering and provides a total configuration at a lower price than if you ordered the same configuration *a la carte*. You can configure additional hardware features, but going below the initial configuration eliminates any price advantage.

If you order an IBM Power 520 (MTM 8203-E4A) server with a 1-, 2-, or 4-core POWER6 4.2 GHz Express Product Offering or a 2- or 4-core POWER6+ 4.7 GHz Express Product Offering as defined here, you might qualify for a processor activation at no additional charge. The number of processors, total memory, quantity and size of disk, and presence of a media device are the only features that determine if a customer is entitled to a processor activation at no additional charge.

When you purchase an Express Product Offering, you are entitled to a lower priced AIX or Linux operating system license, or you can choose to purchase the system with no operating system. The lower priced AIX or Linux operating system is processed using a feature code on AIX and either Red Hat® or SUSE® Linux. You can choose either the lower priced AIX or Linux subscription, but not both. If you choose AIX 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 converse is true if you choose a Linux subscription as your lower priced operating system. Systems with a lower priced AIX offering are referred to as the *IBM Power Systems Express*, *AIX edition* and systems with a lower priced Linux operating system are referred to as the *IBM Power Systems Express, OpenPower® editions*. In the case of Linux, only the first subscription purchased is lower priced. So, for example, additional licenses purchased for Red Hat to run in multiple partitions 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, as well as a discounted AIX or Linux operating system license. However, selection of total memory or DASD that are smaller than the totals defined as the minimum requirements disqualifies the order as an Express Product Offering.

Processor activations are available only to Solution Delivery Integration (SDIs) as MES orders.

1.7.1 1-core 4.2 GHz offering

This section describes the features of the 1-core 4.2 GHz offering.

Deskside configuration

Table 1-11 lists the features and descriptions for the deskside 8203-E4A with only one processor activated.

Table 1-11   8203-E4A deskside configuration (1-core 4.2 GHz) features and descriptions

Feature Code	Description	Qty
1843	Operator Panel Cable, Deskside with 3.5 in. DASD Backplane	1
3647	146.8 GB 15 000 rpm SAS Disk Drive	2
4521	2048 MB (2 x 1024 MB) RDIMMs, 667 MHz, 512 Mb DRAM	1
5623	Dual Port 1 Gb Integrated Virtual Ethernet Daughter Card	1
5633	1-core 4.2 GHz POWER6 Processor Card, 4 Memory DIMM Slots	1
5676	Zero-priced Processor Activation for 5633	1
5743	SATA Slimline DVD-ROM Drive	1
6xxx	Power Cord	2
7112	IBM Deskside Cover Set (With door)	1
7703	Power Supply, 950 watt ac, Hot-swap, Base and Redundant	2
8308	DASD/Media Backplane for 3.5 in. DASD/SATA DVD/Tape (without external SAS port)	1
9300/97xx	Language Group Specify	1

Rack mount configuration

Table 1-12 lists the features and descriptions for the rack mount 8203-E4A with only one processor activated.

Table 1-12   8203-E4A rack mount configuration (1-core 4.2 GHz) features and description

Feature Code	Description	Qty
1877	Op Panel Cable for Rack Mount Drawer with 3.5 in. DASD	1
3647	146.8 GB 15 000 rpm SAS Disk Drive	2
4521	2048 MB (2 x 1024 MB) RDIMMs, 667 MHz, 512 Mb DRAM	2
5623	Dual Port 1 Gb Integrated Virtual Ethernet Daughter Card	1
5633	1-core 4.2 GHz POWER6 Processor Card, 4 Memory DIMM Slots	1
5676	Zero-priced Processor Activation for 5633	1
5743	SATA Slimline DVD-ROM Drive	1
6xxx	Power Cord	2
7114	IBM/OEM Rack Mount Drawer Rail Kit	1
7200	IBM Rack Mount Drawer Bezel and Hardware	1
7703	Power Supply, 950 watt ac, Hot-swap, Base and Redundant	2

Feature Code	Description	Qty
8308	DASD/Media Backplane for 3.5 in. DASD/SATA DVD/Tape (without external SAS port)	1
9300/97xx	Language Group Specify	1

1.7.2 2-core 4.2 GHz offering

This section describes the features of the 2-core 4.2 GHz offering.

Deskside configuration

Table 1-13 lists the features and descriptions for the deskside 8203-E4A with two processors activated.

Table 1-13   8203-E4A deskside configuration (2-core 4.2 GHz) features and descriptions

Feature Code	Description	Qty
1843	Operator Panel Cable, Deskside with 3.5 in. DASD Backplane	1
3647	146.8 GB 15 000 rpm SAS Disk Drive	2
4532	4096 MB (2 x 2048 MB) RDIMMs, 667 MHz, 1 Gb DRAM	1
5623	Dual Port 1 Gb Integrated Virtual Ethernet Daughter Card	1
5634	2-core 4.2 GHz POWER6 Processor Card, 4 Memory DIMM Slots	1
5654	One Processor Activation for Processor Feature 5634	1
5677	Zero-priced Processor Activation for 5634	1
5743	SATA Slimline DVD-ROM Drive	1
6xxx	Power Cord	2
7112	IBM Deskside Cover Set (With door)	1
7703	Power Supply, 950 watt ac, Hot-swap, Base and Redundant	2
8308	DASD/Media Backplane for 3.5 in. DASD/SATA DVD/Tape (without external SAS port)	1
9300/97xx	Language Group Specify	1

Rack mount configuration

Table 1-14 lists the features and descriptions for the rack mount 8203-E4A with two processors activated.

Table 1-14   8203-E4A rack mount configuration (2-core 4.2 GHz) features and descriptions

Feature Code	Description	Qty
1877	Op Panel Cable for Rack Mount Drawer with 3.5 in. DASD	1
3647	146.8 GB 15 000 rpm SAS Disk Drive	2
4532	4096 MB (2 x 2048 MB) RDIMMs, 667 MHz, 1 Gb RDRAM	1
5623	Dual Port 1 Gb Integrated Virtual Ethernet Daughter Card	1
5634	2-core 4.2 GHz POWER6 Processor Card, 4 Memory DIMM Slots	1

Feature Code	Description	Qty
5654	One Processor Activation for Processor Feature 5634	1
5677	Zero-priced Processor Activation for 5634	1
5743	SATA Slimline DVD-ROM Drive	1
6xxx	Power Cord	2
7114	IBM/OEM Rack Mount Drawer Rail Kit	1
7200	IBM Rack Mount Drawer Bezel and Hardware	1
7703	Power Supply, 950 watt ac, Hot-swap, Base and Redundant	2
8308	DASD/Media Backplane for 3.5 in. DASD/SATA DVD/Tape (without external SAS port)	1
9300/97xx	Language Group Specify	1

1.7.3 2-core 4.7 GHz offering

This section describes the features of the 2-core 4.7 GHz offering.

Deskside configuration

Table 1-15 lists the features and descriptions for the deskside 8203-E4A with two processors activated.

Table 1-15   8203-E4A deskside configuration (2-core 4.7 GHz) features and descriptions

Feature Code	Description	Qty
1843	Operator Panel Cable, Deskside with 3.5 in. DASD Backplane	1
3647	146.8 GB 15 000 rpm SAS Disk Drive	2
4532	4096 MB (2 x 2048 MB) RDIMs, 667 MHz, 1 Gb DRAM	1
5577	2-core 4.7 GHz POWER6 Processor Card, 4 Memory DIMM Slots	1
5578	One Processor Activation for Processor Feature 5635 (price per processor)	1
5579	Zero-priced Processor Activation for 5635	1
5623	Dual Port 1 Gb Integrated Virtual Ethernet Daughter Card	1
5743	SATA Slimline DVD-ROM Drive	1
6xxx	Power Cord	2
7112	IBM Deskside Cover Set (With door)	1
7703	Power Supply, 950 watt ac, Hot-swap, Base and Redundant	2
8308	DASD/Media Backplane for 3.5 in. DASD/SATA DVD/Tape (without external SAS port)	1
9300/97xx	Language Group Specify	1

Rack mount configuration

Table 1-16 on page 28 lists the features and descriptions for the rack mount 8203-E4A with two processors activated.

Table 1-16   8203-E4A rack mount configuration (2-core 4.7 GHz) features and descriptions

Feature Code	Description	Qty
1877	Op Panel Cable for Rack Mount Drawer with 3.5 in. DASD	1
3647	146.8 GB 15 000 rpm SAS Disk Drive	2
4532	4096 MB (2 x 2048 MB) RDIMs, 667 MHz, 1 Gb DRAM	1
5577	2-core 4.7 GHz POWER6 Processor Card, 4 Memory DIMM Slots	1
5578	One Processor Activation for Processor Feature 5635 (price per processor)	1
5579	Zero-priced Processor Activation for 5635	1
5623	Dual Port 1 Gb Integrated Virtual Ethernet Daughter Card	1
5743	SATA Slimline DVD-ROM Drive	1
6xxx	Power Cord	2
7114	IBM/OEM Rack-Mount Drawer Rail Kit	1
7703	Power Supply, 950 watt ac, Hot-swap, Base and Redundant	2
8308	DASD/Media Backplane for 3.5 in. DASD/SATA DVD/Tape (without external SAS port)	1
9300/97xx	Language Group Specify	1

1.7.4 4-core 4.7 GHz offering

This section describes the features of the 4-core 4.7 GHz offering.

Deskside configuration

Table 1-17 lists the features and descriptions for the deskside 8203-E4A with all four processors activated.

Table 1-17   8203-E4A deskside configuration (4-core 4.7 GHz) features and descriptions

Feature Code	Description	Qty
1843	Operator Panel Cable, Deskside with 3.5 in. DASD Backplane	1
3647	146.8 GB 15 000 rpm SAS Disk Drive	2
4532	4096 MB (2 x 2048 MB) RDIMs, 667 MHz, 1 Gb DRAM	2
5587	4-core 4.7 GHz POWER6 Processor Card, 8 Memory DIMM Slots	1
5588	One Processor Activation for Processor Feature 5635 (price per processor)	2
5589	Zero-priced Processor Activation for 5635	2
5623	Dual Port 1 Gb Integrated Virtual Ethernet Daughter Card	1
5743	SATA Slimline DVD-ROM Drive	1
6xxx	Power Cord	2
7112	IBM Deskside Cover Set (With door)	1
7703	Power Supply, 950 watt ac, Hot-swap, Base and Redundant	2

Feature Code	Description	Qty
8308	DASD/Media Backplane for 3.5 in. DASD/SATA DVD/Tape (without external SAS port)	1
9300/97xx	Language Group Specify	1

Rack mount configuration

Table 1-18 lists the features and descriptions for the rack mount 8203-E4A with all four processors activated.

Table 1-18   8203-E4A rack mount configuration (4-core 4.7 GHz) features and descriptions

Feature Code	Description	Qty
1877	Op Panel Cable for Rack Mount Drawer with 3.5 in. DASD	1
3647	146.8 GB 15 000 rpm SAS Disk Drive	2
4532	4096 MB (2 x 2048 MB) RDIMs, 667 MHz, 1 Gb DRAM	2
5587	4-core 4.7 GHz POWER6 Processor Card, 4 Memory DIMM Slots	1
5588	One Processor Activation for Processor Feature 5635 (price per processor)	2
5589	Zero-priced Processor Activation for 5635	2
5623	Dual Port 1 Gb Integrated Virtual Ethernet Daughter Card	1
5743	SATA Slimline DVD-ROM Drive	1
6xxx	Power Cord	2
7114	IBM/OEM Rack-Mount Drawer Rail Kit	1
7703	Power Supply, 950 watt ac, Hot-swap, Base and Redundant	2
8308	DASD/Media Backplane for 3.5 in. DASD/SATA DVD/Tape (without external SAS port)	1
9300/97xx	Language Group Specify	1

1.8 IBM Power Systems Solution Editions for AIX / Linux

Designed to take advantage of the combined experience and expertise of IBM and our independent software vendor (ISV) partners, IBM Power Systems Solution Editions provide validated and recommended end-to-end ISV solution stacks running on the Power Systems Express family that are sized and optimized to satisfy a range of user requirements for specific ISV application environments, such as Oracle and SAP.

You can use IBM Power Systems Express configuration-based Solution Editions in conjunction with Oracle or SAP application environments that are tested, validated, and sized through the IBM Oracle and IBM SAP Competency Centers. Together with Oracle and SAP, IBM has developed a proven capacity estimation capability to aid in designing an optimal configuration for each specific client environment. The IBM Power Systems Solution Editions built on the IBM Power Systems Express family take advantage of that experience and expertise to provide a range of choices to assist you in planning for and acquiring Oracle- and SAP-based solutions on IBM Power Systems.

The IBM Power Systems Solution Editions employ the same configuration rules as the Express Product Offerings and are available as an initial configuration option that can be selected using the IBM configuration tool during the configuration of a Power Systems Express server. Prior to order placement, you need to obtain a detailed sizing estimate customized for your environment from the IBM Techline Solution Sizing Team, which is accessible through IBM or your IBM Business Partner representative.

IBM Power Systems Solution Editions support the POWER6 processor-based Power Systems 520 Express as a configuration option.

1.8.1 IBM Power Systems Solution Edition for Oracle

The IBM Power Systems Solution Edition for Oracle includes a set of affordable Express configurations for popular IBM Power Systems Express models that satisfy a range of user requirements from various industries, including finance, manufacturing, and distribution. These Express configurations are tailored to specific user requirements for two Oracle business suites:

▶ Oracle E-Business Suite
▶ JD Edwards® EnterpriseOne

Bringing the Enterprise Resource Planning (ERP) application environment together with the computing infrastructure, the IBM Power Systems Solution Edition for Oracle is an end-to-end ERP solution that is designed to provide a more integrated and tuned environment for the small- or medium-sized business that is challenged with becoming an On Demand Business.

Highlights of this offering include:

▶ Attractively-priced Power Systems Express configuration building blocks tailored to fit popular Oracle E-Business Suite and JD Edwards EnterpriseOne environments

▶ Pre-sized configurations to simplify selection and ease capacity planning for a wide range of user requirements for Oracle E-Business Suite and JD Edwards EnterpriseOne

▶ Support for the diverse needs of users, senior management, and IT professionals

▶ Highly reliable Power Systems Express family designed to help businesses build resilient and scalable computing environments without compromising system affordability

▶ Ability to meet the configuration needs of SMB clients in many industries

Mutual IBM and Oracle clients are eligible to order these new configurations if they are used in conjunction with one of these participating application environments:

▶ **Oracle E-Business Suite** is a fully integrated, comprehensive suite of business applications for enterprises of all sizes that provides functionality to meet your unique requirements. Whether you implement one module at a time, multiple modules, or the complete suite, Oracle E-Business Suite provides better business information for effective decision-making and enables an adaptive enterprise for optimal responsiveness.

▶ **JD Edwards EnterpriseOne** is a complete suite of modular, pre-integrated, industry-specific business applications configured for rapid deployment and ease of administration. The modular design allows you to implement only those solutions your business requires, and the flexible, integrated architecture lowers the ongoing cost of ownership.

The Express configurations are building blocks that can be enhanced with additional memory and I/O options to tailor the infrastructure to satisfy specific client requirements. In addition, an end-to-end solution stack combines the IBM Power Systems Express family of servers,

IBM PowerVM, IBM System Storage, IBM Tivoli® Access Manager, Oracle DataVault, and the powerful Oracle software environments.

Power Systems 520 Oracle Solution Edition 2-core 4.2 GHz

Table 1-19 lists the features and descriptions for the rack mount 8203-E4A with two processors activated.

> **Note:** This configuration is the same as the 8203-E4A 2-core 4.2 GHz Express offering, except 16 GB memory (2 x feature code 4523) replaces 4 GB memory (1 x feature code 4532).

Table 1-19   Power Systems 520 Oracle Solution Edition 2-Core 4.2 GHz configuration

Feature Code	Description	Qty
1877	Op Panel Cable for Rack Mount Drawer with 3.5 in. DASD	1
3647	146.8 GB 15 000 rpm SAS Disk Drive	2
4523	8192 MB (2 x 4096 MB) RDIMMs, 667 MHz, 512 Mb Stacked DRAM	2
5623	Dual Port 1 Gb Integrated Virtual Ethernet Daughter Card	1
5634	2-core 4.2 GHz POWER6 Processor Card, 4 Memory DIMM Slots	1
5654	One Processor Activation for Processor 5634	1
5677	Zero-priced Processor Activation for 5634	1
5743	SATA Slimline DVD-ROM Drive	1
6xxx	Power Cord	2
7114	IBM/OEM Rack Mount Drawer Rail Kit	1
7200	IBM Rack Mount Drawer Bezel and Hardware	1
7703	Power Supply, 950 watt ac, Hot-swap, Base and Redundant	2
8308	DASD/Media Backplane for 3.5 in. DASD/SATA DVD/Tape (without external SAS port)	1
9300/97xx	Language Group Specify	1

Power Systems 520 Oracle Solution Edition 4-Core 4.2 GHz

Table 1-20 lists the features and descriptions for the rack mount 8203-E4A with two processors activated.

Table 1-20   Power Systems 520 Oracle Solution Edition 4-Core -4.2 GHz configuration

Feature Code	Description	Qty
1877	Op Panel Cable for Rack Mount Drawer with 3.5 in. DASD	1
3647	146.8 GB 15 000 rpm SAS Disk Drive	2
4523	8192 MB (2 x 4096 MB) RDIMMs, 667 MHz, 512 Mb Stacked DRAM	4
5623	Dual Port 1 Gb Integrated Virtual Ethernet Daughter Card	1

Feature Code	Description	Qty
5635	4-core 4.2 GHz POWER6 Processor Card, 8 Memory DIMM Slots	1
5654	One Processor Activation for Processor 5634	2
5677	Zero-priced Processor Activation for 5634	2
5743	SATA Slimline DVD-ROM Drive	1
6xxx	Power Cord	2
7114	IBM/OEM Rack Mount Drawer Rail Kit	1
7200	IBM Rack Mount Drawer Bezel and Hardware	1
7703	Power Supply, 950 watt ac, Hot-swap, Base and Redundant	2
8308	DASD/Media Backplane for 3.5 in. DASD/SATA DVD/Tape (without external SAS port)	1
9300/97xx	Language Group Specify	1

1.8.2 IBM Power Systems Solution Edition for SAP

The IBM Power Systems Solution Edition for SAP offering helps reduce the effort implementing an SAP environment by providing pre-sized hardware configurations based on the IBM Power Systems Express family, running a proven, comprehensive suite of modular, pre-integrated business software solutions based on SAP ERP or SAP NetWeaver® BI.

IBM Power Systems Solution Edition for SAP includes affordable Power Systems Express configurations recommended by the IBM SAP International Competence Center that satisfy a range of user sizing requirements for running the popular SAP ERP and SAP NetWeaver BI applications. The tailored IBM Power Systems Express configurations are affordable server building blocks that can be enhanced with additional memory and I/O options to tailor the infrastructure to satisfy specific user requirements.

Mutual IBM and SAP clients are eligible to order these new configurations if they are used in conjunction with one of these participating application environments:

► SAP ERP consists of a set of proven, robust solutions with extended cross-enterprise function. As a result, corporate assets and critical business processes can be better managed. Today, more than ever, companies must have a competitive advantage to prosper and grow in the ever-changing and highly competitive SMB marketplace. New competitors, market innovations, and better informed customers increase the demands on their business and employees. To succeed in this environment, they need to provide managers and employees with the right tools, applications, and information, and all in real time. SAP ERP helps manage IT costs more effectively by protecting and leveraging the investments already made. Built on the SAP NetWeaver platform, mySAP™ ERP includes the following individual solutions that support key functional areas:

  – mySAP ERP Financials
  – mySAP ERP Human Capital Management
  – mySAP ERP Operations
  – mySAP ERP Corporate Services

► SAP NetWeaver BI provides companies of every size with a highly reliable, high performance business intelligence (BI) solution that can scale to meet their needs. In today's business environment, it is critical to turn the reams of data collected into usable

real-time analysis and reports that can be used to make daily business decisions. Gone are the days when you can wait overnight for reports to be generated. Companies need a BI solution that can analyze their data in a heterogeneous environment and provide the results directly to other business applications, management, or individual users, all in real time. Not only does the BI solution need to be fast, but it also needs to be reliable and scale as a company grows. SAP NetWeaver BI provides a high performance, scalable BI solution, delivering real-time business analytics to help clients make daily business decisions. It provides a highly reliable and scalable platform that can grow with the BI client's business without compromising system affordability.

Power Systems 520 SAP Solution Edition, 2-core 4.2 GHz

Table 1-21 lists the features and descriptions for the rack mount 8203-E4A with two processors activated.

Note: This configuration is the same as the 8203-E4A 2-core 4.2 GHz Express offering, except 16 GB memory (2 x feature code 4523) replaces 4 GB memory (1 x feature code 4532).

Table 1-21   Power Systems 520 SAP Solution Edition (2-core 4.2 GHz) rack mount configuration

Feature Code	Description	Qty
1877	Op Panel Cable for Rack Mount Drawer with 3.5 in. DASD	1
3647	146.8 GB 15 000 rpm SAS Disk Drive	2
4523	8192 MB (2 x 4096 MB) RDIMMs, 667 MHz, 512 Mb Stacked DRAM	2
5623	Dual Port 1 Gb Integrated Virtual Ethernet Daughter Card	1
5634	2-core 4.2 GHz POWER6 Processor Card, 4 Memory DIMM Slots	1
5654	One Processor Activation for Processor 5634	1
5677	Zero-priced Processor Activation for 5634	1
5743	SATA Slimline DVD-ROM Drive	1
6xxx	Power Cord	2
7114	IBM/OEM Rack Mount Drawer Rail Kit	1
7200	IBM Rack Mount Drawer Bezel and Hardware	1
7703	Power Supply, 950 watt ac, Hot-swap, Base and Redundant	2
8308	DASD/Media Backplane for 3.5 in. DASD/SATA DVD/Tape (without external SAS port)	1
9300/97xx	Language Group Specify	1

Power Systems 520 SAP Solution Edition 4-core 4.2 GHz

Table 1-22 lists the features and descriptions for the rack mount 8203-E4A with all four processors activated.

Table 1-22   Power Systems 520 SAP Solution Edition (4-core 4.2 GHz) rack mount configuration

Feature Code	Description	Qty
1877	Op Panel Cable for Rack Mount Drawer with 3.5 in. DASD	1
3647	146.8 GB 15 000 rpm SAS Disk Drive	2
4523	8192 MB (2 x 4096 MB) RDIMMs, 667 MHz, 512 Mb Stacked DRAM	4
5623	Dual Port 1 Gb Integrated Virtual Ethernet Daughter Card	1
5635	4-core 4.2 GHz POWER6 Processor Card, 8 Memory DIMM Slots	1
5655	One Processor Activation for Processor 5635	2
5678	Zero-priced Processor Activation for 5635	2
5743	SATA Slimline DVD-ROM Drive	1
6xxx	Power Cord	2
7114	IBM/OEM Rack Mount Drawer Rail Kit	1
7200	IBM Rack Mount Drawer Bezel and Hardware	1
7703	Power Supply, 950 watt ac, Hot-swap, Base and Redundant	2
8308	DASD/Media Backplane for 3.5 in. DASD/DVD/Tape (without external SAS port)	1
9300/97xx	Language Group Specify	1

1.9 IBM Power Systems Express Editions for IBM i

IBM i initially offered POWER6 Express Editions and Solution Editions on the 9407-M15 and 9408-M25 MTMs. Effective with 8203-E4A system firmware level EL340_039, these Express Edition and Solution Edition "initial configurations" are also supported on the 8203-E4A for 1-core, 2-core and 4-core configurations.

These IBM i Express Editions and Solution Editions enable initial ease of ordering and feature a lower price than if you order them *build-to-order* or *a la carte*. Taking advantage of the edition is the only way that you can use no-charge features for processor activations and IBM i user license entitlements.

The Express Editions are available only during the initial system order and cannot be ordered after your system ships.

The IBM configurator offers these easy-to-order Express Editions that include no-charge activations or no-charge IBM i user entitlements. You can modify the Express Edition configurations to match your exact requirements for your initial shipment, increasing or decreasing the configuration. However, If you create a configuration that falls below any of the defined minimum requirements, the IBM configurator replaces the no-charge features with equivalent function regular charge features.

A Solution Edition on IBM i is available for SAP application users for the 2-core 8203-E4A. Users of SAP's mySAP, ERP, BI, CRM, PLM, and SCM can qualify to use this edition.

A Solution Edition on IBM i for JD Edwards EnterpriseOne and World application users is available for the 2-core 8203-E4A.

The Power 520 Solution Editions for SAP applications require proof of a minimum purchase before the system ships from IBM. For details, visit the Solution Edition Web site at:

http://www.ibm.com/systems/i/editions/solutionedition.html

For the validation form and entitlement requirements, visit:

http://www.ibm.com/systems/i/editions/validate.html

Power Systems 520 Entry Express Edition, 1-core (#9633)

To use the no-charge features on your initial order, you must order a minimum of:

– One IBM i processor license
– Five IBM i user entitlements
– 2 GB memory
– Two SAS IBM i disk drives (mirroring assumed)

For the lower price, one no-charge processor activation and five no-charge IBM i user entitlements are available. Additionally, System i Access unlimited users is available at no charge.

The 1-core Entry Edition (#9633) suggested starting configuration is:

► One 1-core 4.2 GHz processor card (#5633)
► One 2 GB memory feature (#4521)
► Two 139.5 GB 15k rpm disk drives (#3677)
► One dual-port 1 Gb integrated Ethernet adapter (#5623)
► Two power supplies, 950 watt, base (#7703)
► One PCIe WAN IOA (#2893 or #2894) (country dependent)
► One DVD-ROM (#5743)
► One DASD/media backplane (#8310)
► One 80/160 GB DAT160 SAS tape drive (#5619)
► Two power cords (6xxx)
► One processor activation (#5676) (no additional charge)
► Five IBM i user entitlements (no additional charge) (IBM i 5.4 5722-SSC or IBM i 6.1 5761-SSC)
► Five IBM i user entitlements (charge) (IBM i 5.4 5722-SSC or IBM i 6.1 5761-SSC)
► System i Access unlimited users (5722-XW1 or 5761-XW1)
► WebSphere® Development Studio and Rational® development software (V5R4: 5722-WDS, V6R1: 5761-WDS and 5733-RDI)
► PowerVM Standard Edition (#8506), or higher
► One year of software maintenance

Power Systems 520 Growth Express Edition, 1-core (#9634)

To use the no-charge features on your initial order, you must order a minimum of:

– One IBM i processor license
– Ten IBM i user entitlements
– Four GB memory
– Four SAS IBM i disk drives (RAID-5 or mirroring assumed)

For the lower price, one no-charge processor activation and ten no-charge IBM i user entitlements are available.

1-core Growth Edition (#9634) suggested starting configuration:

► One 1-core 4.2 GHz processor card (#5633)
► One 4 GB memory feature (#4522)
► Four 139.5 GB 15k rpm disk drives (#3677) (RAID-5 or mirroring assumed)

- ► One SAS RAID Enablement (#5679)
- ► One quad-port 1 Gb integrated Ethernet adapter (#5624)
- ► Two power supplies, 950 watt, base (#7703)
- ► One PCIe WAN IOA (#2893 or #2894) (country dependent)
- ► One DVD-ROM (#5743)
- ► One DASD/media backplane (#8310)
- ► One 80/160 GB DAT160 SAS tape drive (#5619)
- ► Two power cords (6xxx)
- ► One processor activation (#5676) (no additional charge)
- ► Ten IBM i user entitlements (no additional charge) (IBM i 5.4 5722-SSC or IBM i 6.1 5761-SSC)
- ► Ten IBM i user entitlements (charge) (IBM i 5.4 5722-SSC or IBM i 6.1 5761-SSC)
- ► System i Access unlimited users (5722-XW1/5761-XW1)
- ► WebSphere Development Studio and Rational development software (V5R4: 5722-WDS, V6R1: 5761-WDS and 5733-RDI)
- ► PowerVM Standard Edition (#8506), or higher
- ► DB2® Web Query (5733-QU2)
- ► One year of software maintenance

Power Systems 520 30 User Express Edition, 2-core (#9636)

To receive a no-charge feature on the initial order, you must order a minimum of:

- – One IBM i processor license
- – Thirty unlimited IBM i user entitlements for the 2-core or 50 entitlements for the 4-core 8203
- – 4 GB memory
- – Four SAS or SCSI IBM i disk drives (any size)
- – One SAS RAID Enablement (#5679) for the system unit or one 1.5 GB write cache disk controller (#5583/5782/5778/5904/5908)
- – Two power supplies (2 x #7703)
- – PowerVM Standard Edition (#8506), or higher

The no-charge feature is two no-charge processor activations. Depending on IBM country-specific definitions, a services voucher may also be available to you.

The suggested starting configuration includes:

- ► One 2-core 4.7 GHz processor card (#5577)
- ► Two processor activations feature 5579 at no charge
- ► 30/50/150/unlimited IBM i user entitlements, depending on the edition
- ► One 8 GB memory feature (#4523)
- ► Six 139.5 GB 15k rpm SAS disk drives (#3677)
- ► One SAS RAID Enablement (#5679)
- ► One Quad-port 1 Gb integrated Ethernet adapter (#5624)
- ► Two power supplies, 950 watt (#7703)
- ► One PCIe WAN IOA (#2893 or 2894) (country dependent)
- ► One SATA DVD-ROM (#5743)
- ► One DASD/media backplane with external SAS port (#8310)
- ► One 80/160 GB DAT160 SAS tape drive (#5619)
- ► Two power cords (#6xxx)
- ► One GX adapter for 12X I/O loop
- ► One 9910 UPS
- ► One UPS communication cable (#1827)
- ► One IBM i processor entitlement (chargeable) (57xx–SSA)

Power Systems 520 150 User Express Edition, 2-core (#9637)

To receive a no-charge feature on the initial order, you must order a minimum of:

– One IBM i processor license
– One hundred fifty IBM i user entitlements
– 4 GB memory
– Four SAS or SCSI IBM i disk drives (any size)
– One SAS RAID Enablement (#5679) for the system unit or one 1.5 GB write cache disk controller (#5583/5782/5778/5904/5908)
– Two power supplies (2 x #7703)
– PowerVM Standard Edition (#8506), or higher

The no-charge feature is two no-charge processor activations. Depending on IBM country-specific definitions, a services voucher may also be available to you.

The suggested starting configuration includes:

► One 2-core 4.7 GHz processor card (#5577)
► Two processor activations feature 5579 at no charge
► 30/50/150/unlimited IBM i user entitlements, depending on the edition
► One 8 GB memory feature (#4523)
► Six 139.5 GB 15k rpm SAS disk drives (#3677)
► One SAS RAID Enablement (#5679)
► One Quad-port 1 Gb integrated Ethernet adapter (#5624)
► Two power supplies, 950 watt (#7703)
► One PCIe WAN IOA (#2893 or 2894) (country dependent)
► One SATA DVD-ROM (#5743)
► One DASD/media backplane with external SAS port (#8310)
► One 80/160 GB DAT160 SAS tape drive (#5619)
► Two power cords (#6xxx)
► One GX adapter for 12X I/O loop
► One 9910 UPS
► One UPS communication cable (#1827)
► One IBM i processor entitlement (chargeable) (57xx--SSA)

Power Systems 520 Unlimited User Express Edition, 2-core (#9638)

To receive a no-charge feature on the initial order, you must order a minimum of:

– One IBM i processor license
– Unlimited IBM i user entitlements
– 4 GB memory
– Four SAS or SCSI IBM i disk drives (any size)
– One SAS RAID Enablement (#5679) for the system unit or one 1.5 GB write cache disk controller (#5583/5782/5778/5904/5908)
– Two power supplies (2 x #7703)
– PowerVM Standard Edition (#8506), or higher

The no-charge feature is two no-charge processor activations. Depending on IBM country-specific definitions, a services voucher may also be available to you.

The suggested starting configuration includes:

► One 2-core 4.7 GHz processor card (#5577)
► Two processor activations feature 5579 at no charge
► 30/50/150/unlimited IBM i user entitlements, depending on the edition
► One 8 GB memory feature (#4523)
► Six 139.5 GB 15k rpm SAS disk drives (#3677)
► One SAS RAID Enablement (#5679)

- ► One Quad-port 1 Gb integrated Ethernet adapter (#5624)
- ► Two power supplies, 950 watt (#7703)
- ► One PCIe WAN IOA (#2893 or 2894) (country dependent)
- ► One SATA DVD-ROM (#5743)
- ► One DASD/media backplane with external SAS port (#8310)
- ► One 80/160 GB DAT160 SAS tape drive (#5619)
- ► Two power cords (#6xxx)
- ► One GX adapter for 12X I/O loop
- ► One 9910 UPS
- ► One UPS communication cable (#1827)
- ► One IBM i processor entitlement (chargeable) (57xx--SSA)

Power Systems 520 50 User Express Edition, 4-core (#9639)

To receive a no-charge feature on the initial order, you must order a minimum of:

- – One IBM i processor license
- – Thirty unlimited IBM i user entitlements for the 2-core or 50 entitlements for the 4-core 8203
- – 4 GB memory
- – Four SAS or SCSI IBM i disk drives (any size)
- – One SAS RAID Enablement (#5679) for the system unit or one 1.5 GB write cache disk controller (#5583/5782/5778/5904/5908)
- – Two power supplies (2 x #7703)
- – PowerVM Standard Edition (#8506), or higher

The no-charge feature is two no-charge processor activations. Depending on IBM country-specific definitions, a services voucher may also be available to you.

The suggested starting configuration includes:

- ► One 4-core 4.7 GHz processor card (#5587)
- ► Two processor activations feature 5579 at no charge
- ► 30/50/150/unlimited IBM i user entitlements, depending on the edition
- ► One 8 GB memory feature (#4523)
- ► Six 139.5 GB 15k rpm SAS disk drives (#3677)
- ► One SAS RAID Enablement (#5679)
- ► One Quad-port 1 Gb integrated Ethernet adapter (#5624)
- ► Two power supplies, 950 watt (#7703)
- ► One PCIe WAN IOA (#2893 or 2894) (country dependent)
- ► One SATA DVD-ROM (#5743)
- ► One DASD/media backplane with external SAS port (#8310)
- ► One 80/160 GB DAT160 SAS tape drive (#5619)
- ► Two power cords (#6xxx)
- ► One GX adapter for 12X I/O loop
- ► One 9910 UPS
- ► One UPS communication cable (#1827)
- ► One IBM i processor entitlement (chargeable) (57xx--SSA)
- ► PowerVM Standard Edition (#8506), or higher
- ► System i Access unlimited users (57xx-XW1)
- ► WebSphere Development Studio and Rational development software (V5R4: 5722-WDS, V6R1: 5761-WDS and 5733-SOA)
- ► DB2 Web Query (5733-QU2)
- ► One year of software maintenance

Power Systems 520 150 User Express Edition, 4-core (#9640)

To receive a no-charge feature on the initial order, you must order a minimum of:

– One IBM i processor license
– One hundred fifty IBM i user entitlements
– 4 GB memory
– Four SAS or SCSI IBM i disk drives (any size)
– One SAS RAID Enablement (#5679) for the system unit or one 1.5 GB write cache disk controller (#5583/5782/5778/5904/5908)
– Two power supplies (2 x #7703)
– PowerVM Standard Edition (#8506), or higher

The no-charge feature is two no-charge processor activations. Depending on IBM country-specific definitions, a services voucher may also be available to you.

The suggested starting configuration includes:

► One 4-core 4.7 GHz processor card (#5587)
► Two processor activations feature 5579 at no charge
► 30/50/150/unlimited IBM i user entitlements, depending on the edition
► One 8 GB memory feature (#4523)
► Six 139.5 GB 15k rpm SAS disk drives (#3677)
► One SAS RAID Enablement (#5679)
► One Quad-port 1 Gb integrated Ethernet adapter (#5624)
► Two power supplies, 950 watt (#7703)
► One PCIe WAN IOA (#2893 or 2894) (country dependent)
► One SATA DVD-ROM (#5743)
► One DASD/media backplane with external SAS port (#8310)
► One 80/160 GB DAT160 SAS tape drive (#5619)
► Two power cords (#6xxx)
► One GX adapter for 12X I/O loop
► One 9910 UPS
► One UPS communication cable (#1827)
► One IBM i processor entitlement (chargeable) (57xx--SSA)
► PowerVM Standard Edition (#8506), or higher
► System i Access unlimited users (57xx-XW1)
► WebSphere Development Studio and Rational development software (V5R4: 5722-WDS, V6R1: 5761-WDS and 5733-SOA)
► DB2 Web Query (5733-QU2)
► One year of software maintenance

Power Systems 520 Unlimited Express Edition, 4-core (#9643)

To receive a no-charge feature on the initial order, you must order a minimum of:

– One IBM i processor license
– Unlimited IBM i user entitlements
– 4 GB memory
– Four SAS or SCSI IBM i disk drives (any size)
– One SAS RAID Enablement (#5679) for the system unit or one 1.5 GB write cache disk controller (#5583/5782/5778/5904/5908)
– Two power supplies (2 x #7703)
– PowerVM Standard Edition (#8506), or higher

The no-charge feature is two no-charge processor activations. Depending on IBM country-specific definitions, a services voucher may also be available to you.

The suggested starting configuration includes:

- One 4-core 4.7 GHz processor card (#5587)
- Two processor activations feature 5579 at no charge
- 30/50/150/unlimited IBM i user entitlements, depending on the edition
- One 8 GB memory feature (#4523)
- Six 139.5 GB 15k rpm SAS disk drives (#3677)
- One SAS RAID Enablement (#5679)
- One Quad-port 1 Gb integrated Ethernet adapter (#5624)
- Two power supplies, 950 watt (#7703)
- One PCIe WAN IOA (#2893 or 2894) (country dependent)
- One SATA DVD-ROM (#5743)
- One DASD/media backplane with external SAS port (#8310)
- One 80/160 GB DAT160 SAS tape drive (#5619)
- Two power cords (#6xxx)
- One GX adapter for 12X I/O loop
- One 9910 UPS
- One UPS communication cable (#1827)
- One IBM i processor entitlement (chargeable) (57xx--SSA)
- PowerVM Standard Edition (#8506), or higher
- System i Access unlimited users (57xx-XW1)
- WebSphere Development Studio and Rational development software (V5R4: 5722-WDS, V6R1: 5761-WDS and 5733-SOA)
- DB2 Web Query (5733-QU2)
- One year of software maintenance

1.10 IBM Power Systems Solution Editions for IBM i

IBM i on the 8203-E4A offers a Solution Edition that helps meet the needs of Oracle JD Edwards and SAP application users. Users of JD Edwards EnterpriseOne and World applications and SAP mySAP, ERP, BI, CRM, PLM, and SCM can qualify to use this edition. The IBM Power 520 Solution editions for SAP and Oracle applications require proof of a minimum purchase before the system ships from IBM. For details, visit the Solution Edition Web site at:

http://www.ibm.com/systems/i/editions/solutionedition.html

For the validation form and entitlement requirements, visit:

http://www.ibm.com/systems/i/editions/validate.html

When you purchase a new 2-core Power 520 system with a 30/50/150/unlimited i Edition, you may be entitled to receive one or two service vouchers at no additional charge, depending upon IBM country-specific definitions. Systems ordered without an edition feature are not eligible.

Upgrades from a 9406-520/525 keeping the same serial number do not use an edition feature and therefore are not eligible. Solution editions are not eligible. Service vouchers deliver the technical leadership and consulting resources that can help you more fully understand and use the latest features and capabilities of the IBM i operating system. The experts will join your team and help get you started on the road to success with your new operating environment. For more information about vouchers, visit:

http://www.ibm.com/systems/i/hardware/editions/vouchers.html

IBM Power 520 Solution Edition (9635) includes additional no-charge IBM i user license entitlements over the regular Express Editions, resulting in a lower initial list price for qualifying clients.

The no-charge features are two no-charge processor activations and 10 no-charge IBM i user entitlements.

To receive no-charge features on your initial Solution Edition order, your initial order must include a minimum of:

► One IBM i processor license
► Thirty IBM i user entitlements
► 4 GB memory
► Four SAS or SCSI IBM i disk drives (any size)
► SAS RAID Enablement (#5679) for the system unit or a 1.5 GB write cache disk controller (#5583/5782/5778/5904/5908)
► Two power supplies (#7703)

Power Systems 520 Solution Edition, 2-core 4.7 GHz

Table 1-23 lists the suggested features and descriptions for the rack mount 8203-E4A with two processors activated.

Table 1-23   Power Systems 520 SAP Solution Edition (2-core 4.7 GHz) rack mount configuration

Feature Code	Description	Qty
1877	Op Panel Cable for Rack Mount Drawer with 3.5 in. DASD	1
2894	PCIe 2-Line WAN w/Modem CIM	1
3677	139.5 GB 15 000 rpm SAS Disk Drive	6
4523	8192 MB (2 x 4096 MB) RDIMMs, 667 MHz, 512 Mb Stacked DRAM	1
5616	GX Dual-port 12x Channel Attach	1
5624	4-Port 1 Gb Integrated Virtual Ethernet Daughter Card	1
5577	2-core 4.7 GHz POWER6 Processor Card, 4 Memory DIMM Slots	1
5579	Zero-priced Processor Activation for 5577	2
5679	SAS RAID Enablement	1
5743	SATA Slimline DVD-ROM Drive	1
6xxx	Power Cord	2
7114	IBM/OEM Rack Mount Drawer Rail Kit	1
7200	IBM Rack Mount Drawer Bezel and Hardware	1
7703	Power Supply, 950 watt ac, Hot-swap, Base and Redundant	2
8310	DASD/Media Backplane for 3.5 in. DASD/SATA DVD/Tape (with external SAS port)	1
9300/97xx	Language Group Specify	1
1816	IBM i per Processor Registration (E4A)	1
1843	OTC per 10 Users (E4A 2/4-Core)	1

1.11 Upgrades

MES upgrades into the 9408-M25 and the 8203-E4A from the 9406-525 and from selected 9406-520 editions are available. They preserve the existing machine serial number. IBM i processor and user license entitlements are also preserved during the upgrade.

Effective November, 2008, upgrades from the 9408-M25 to the 8203-E4A are supported.

1.12 System racks

In this section, we list the supported racks for the IBM POWER6 technology Power 520 server.

> **Note:** When using this information, keep in mind the following:
>
> ► Where appropriate, we list the System i feature first, followed by the System p feature. In some cases, the separate features might be supported only in upgraded configurations (and not new configurations).
>
> ► In cases where a physical rack is the same for both a System i and a System p order number and that rack is supported on POWER6, the IBM configurator tool might list the System i or System p rack number, depending on the primary operating system specified on the order. While physically the same rack, some supported features within the rack might be listed by default based upon the primary operating system specified.
>
> Some features that are delivered as part a either a System p rack or a System i rack might be different. For example, be sure to order the FC#0553 only when required to support rack integration of MES orders prior to shipment from IBM manufacturing.

Table 1-24 lists the supported racks for the IBM Power System technology.

Table 1-24   Supported racks for the IBM Power System technology

Model	Feature	Description
Power 520, Power 550, servers	0551 and 7014-T00	19 in. 1.8 m 36U Rack. 0551 is equivalent to the 7014-T00 rack.
	0553 and 7014-T42	19 in. 2.0 m 42U Rack. 0553 is equivalent to the 7014-T42 rack.
	0554 and 7014-S11	19 in. 0.6 m 11 U Rack. 0554 is equivalent to the 7014-S11 rack. 0554 or cannot be ordered new as of April 2008.
	0555 and 7014-S25	19 in. 1.3 m (25U). 0555 is equivalent to the 7014-S25 rack.
	7014-B42	19 in. 2.0m 42U Rack. 7014-B42 is originally designed for IBM Blade-Center® H and is available for IBM POWER technology servers also as of February 2008.

1.13 AC power distribution unit and rack content

Power distribution units (PDUs) are required with 7014-T00, 7014-T42 IBM racks and are optional with the 7014-B42 rack. PDUs are optional with the 0551, 0553, and 0555 IBM racks, except when a 0578 or 0588 expansion unit is being ordered.

For rack models T00 and T42, 12-outlet PDUs are available. These include PDUs Universal UTG0247 Connector (#9188 and #7188) and Intelligent PDU+ Universal UTG0247 Connector (#5889 and #7109). For rack Models S11 and S25, #7188 and #7109 are available.

Four PDUs can be mounted vertically in the back of the T00 and T42 racks. See Figure 1-7 on page 43 for the placement of the four vertically mounted PDUs. In the rear of the rack, two additional PDUs can be installed horizontally in the T00 rack and three in the T42 rack. The four vertical mounting locations 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 are available rack spaces.
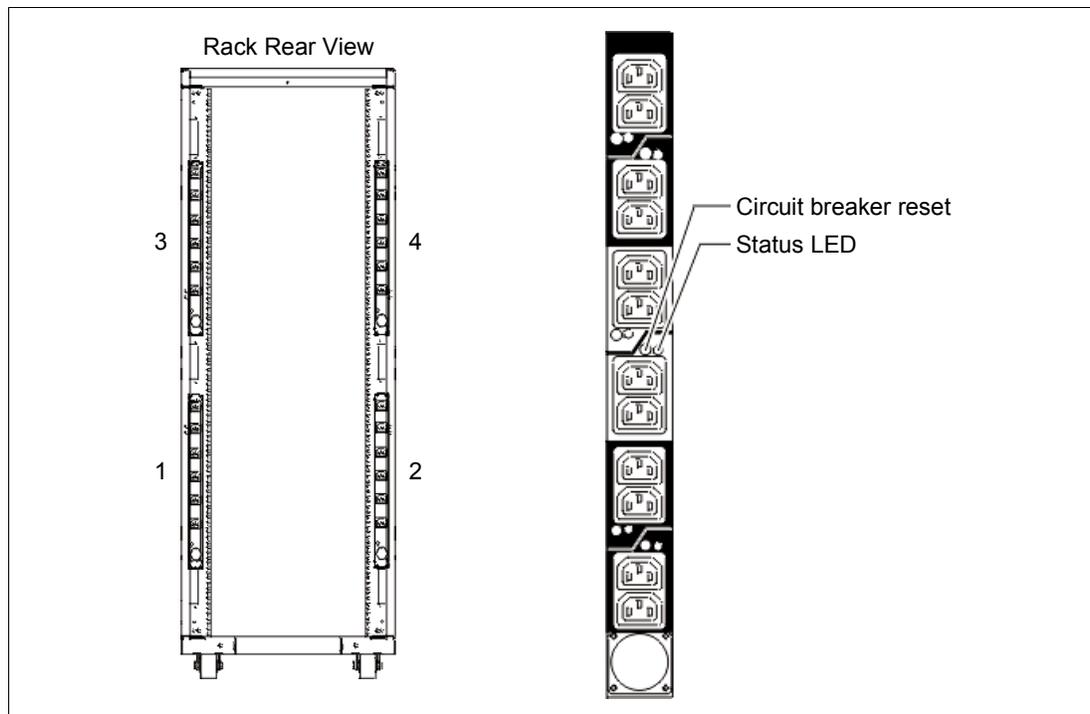
Figure 1-7   PDU placement and PDU view

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

The Base/Side Mount Universal PDU (#9188) and the optional, additional, Universal PDU (#7188) and the Intelligent PDU+ options (#5889 and #7109) 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. Various 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.

Note: For maximum availability it is highly recommended to connect power cords from the same system to two separate PDUs in the rack, and to connect each PDU to an independent power source.

One of the following line cords must be used to distribute power from a wall outlet to the #7188:

► #6489 - 14-Ft 3PH/24A Power Cord
► #6491 - 14-Ft 1PH/63A Pwr Cord
► #6492 - 14-Ft 1PH/48-60A Pwr Cord
► #6653 - 14-Ft 3PH/16A Power Cord
► #6654 - 14-Ft 1PH/24-30A Pwr Cord
► #6655 - 14-Ft 1PH/24-30A WR Pwr Cord
► #6656 - 14-Ft 1PH/32A Power Cord
► #6657 - 14-Ft 1PH/24A Power Cord
► #6658 - 14-Ft 1PH/24A Pwr Cd-Korea

The PDU is Customer Set Up (CSU)

1.13.1 Power Distribution Unit Plus (PDU+)

Energy consumption is becoming a large issue in computer-based businesses. The energy required to power and cool computers can be a significant cost to a business – reducing profit margins and consuming resources.

For all systems without an internal thermal and power consumption method the IBM Power Distribution Unit plus (PDU+) provides a solution to measure and monitor power data. This PDU+ is also sometimes referred to as the intelligent PDU (iPDU), but in this publication we refer to it as the PDU+.

The PDU+ is an intelligent Alternating Current (AC) power distribution unit that monitors the amount of power being used by the devices that are plugged into it. The PDU+ provides twelve C13 power outlets and receives power through a Souriau UTG0247 connector. The PDU+ can be used in many geographies and for many applications by varying the PDU-to-wall power cord, which must be ordered separately (available cord feature numbers listed later). Each PDU+ requires one PDU-to-wall power cord. When the PDU+ is connected to a dedicated power source, it conforms to UL60950, CSA C22.2-60950, EN-60950, and IEC-60950 standards.

There are several feature codes associated with the PDU+:

► 5889 - First PDU+ (with 12 IEC 320-C13 outlets) installed on a 7014-S25, 0555, 7014-T00, 0551, 7014-T42, or 0553 rack

► 7109 - Additional PDU+ (with 12 IEC 320-C13 outlets) installed on a 7014-S25, 0555, 7014-T00, 0551, 7014-T42, or 0553 rack

► 7189 - PDU+ (with six IEC 320-C19 outlets) installed on a 7014-B42 rack

► 7196 - PDU+ (with six IEC320-C19 outlets with an attached power cord) installed on a 7014-B42 rack

Supported power cords include the following:

► #6489 - 14-Ft 3PH/24A Power Cord
► #6491 - 14-Ft 1PH/63A Pwr Cord
► #6492 - 14-Ft 1PH/48-60A Pwr Cord
► #6653 - 14-Ft 3PH/16A Power Cord
► #6654 - 14-Ft 1PH/24-30A Pwr Cord
► #6655 - 14-Ft 1PH/24-30A WR Pwr Cord
► #6656 - 14-Ft 1PH/32A Power Cord
► #6657 - 14-Ft 1PH/24A Power Cord
► #6658 - 14-Ft 1PH/24A Pwr Cd-Korea

The PDU is Customer Set Up (CSU).

Table 1-25 summarizes the physical specifications of the PDU+.

Table 1-25   PDU specifications

Specification	Value or description
Height	43.9 mm (1.73 in.)
Width	447 mm (17.6 in.)
Depth	350 mm (13.78 in.)
Additional Clearance	25 mm (0.98 in.) for circuit breakers 3 mm (0.12 in.) for outlets
Weight (not including power cord)	6.3 kg (13.8 lb.)
Weight of power cord (approximate)	5.4 kg (11.8 lb.)
Operating temperature at 0 - 914 m (0 - 3000 ft) (room ambient)	10 - 35 degrees C (50 - 95 degrees F)
Operating temperature at 914 - 2133 m (3000 - 7000 ft) (room ambient)	10 - 32 degrees C (50 - 90 degrees F)
Operating humidity	8 - 80% (noncondensing)
Localized air temperature in PDU	60 degrees C (140 degrees F) maximum
Rated voltage for 5889, 7109 and 7189 PDU+	220 - 240 or 380 - 415 V, three phase, Wye, 32 A 220 - 240 or 380 - 415 V, three phase, Wye, 16A 220 - 240 V, single phase, 32 A 220 - 240 V, single phase, 63 A 200 - 208 V, single phase, 24 A 200 - 208 V, single phase, 48 A
Rated voltage for 7196 PDU+	200 - 208 V, three phase, Delta, 48 A

Specification	Value or description
Rated frequency (all feature codes)	50 Hz - 60 Hz
Circuit breakers	Six double-pole branch rated circuit breakers rated at 20 A
Power cord features for 5889 and 7109 PDU+	6489, 6491, 6492, 6653, 6654, 6655, 6656, 6657, and 6658
Power cord features for 7189 PDU+	6489, 6491, 6492, 6653
Power cord features for 7196 PDU+	Fixed power cord with IEC 309, 3P+G, 60 A plug
Power outlets for feature codes 7189 and 7196	Six IEC 320-C19 outlets rated at 16 A (VDE) or 20 A (UL/CSA)
Power outlets for feature codes 5889 and 7109	12 IEC 320-C13 outlets rated at 10 A (VDE) or 15 A (UL/CSA)
RS232 serial connector	For firmware update
RJ-45 console connector	Provides a connection using a DB9 to RJ45 provided cable to a notebook computer as a configuration console
RJ-45 Ethernet (LAN) connector	Port to configure the PDU for communication to an Active Energy Manager workstation

You can find out more about the PDU+ and its power-monitoring capabilities at the IBM Systems Hardware Information Center at:

http://publib.boulder.ibm.com/infocenter/systems/scope/hw/index.jsp

Search with "7109" or "PDF+" and follow the links.

Use the following URL to go directly to setting up power monitoring using the PDU+:

http://publib.boulder.ibm.com/infocenter/systems/scope/hw/topic/iphbf/pdupowermonitor.htm

The PDU+ can communicate its power monitor information using its Ethernet connection to a PC workstation running the IBM Active Energy Manager extension to the well-known IBM Systems Director, release level 5.20.2.

IBM Systems Director Active Energy Manager V3.1 is the next generation product of IBM PowerExecutive™, which was previously available for IBM x86 architecture systems only. Active Energy Manager now supports additional IBM platforms and provides new capabilities that build upon the functions previously available with IBM PowerExecutive V2.1.

Active Energy Manager, as an extension of IBM Systems Director 5.20.2, is available for installation on Linux, Linux on x86 architecture, and Microsoft® Windows® to help you monitor and manage the power usage of systems across IBM systems and non-IBM systems.

Originally designed to support IBM BladeCenter and System x®, Active Energy Manager now supports the power management of additional IBM servers, including POWER6 processor-based systems, as well as storage devices using the iPDU.

Administrators at the Active Energy Manager console can use the information to monitor the power consumption of an entire data center at various times of the day, week, or month. They can use that information to identify anomalies and manage loads when electrical demands or

costs are high. Electrical usage information correlates to heat output, so the data can also offer insights into air conditioning requirements.

When a properly configured PDU+ is selected to be managed by the Active Energy Manager it appears on its console display. Figure 1-8 on page 47 shows an example of how a PDU+ can be viewed.
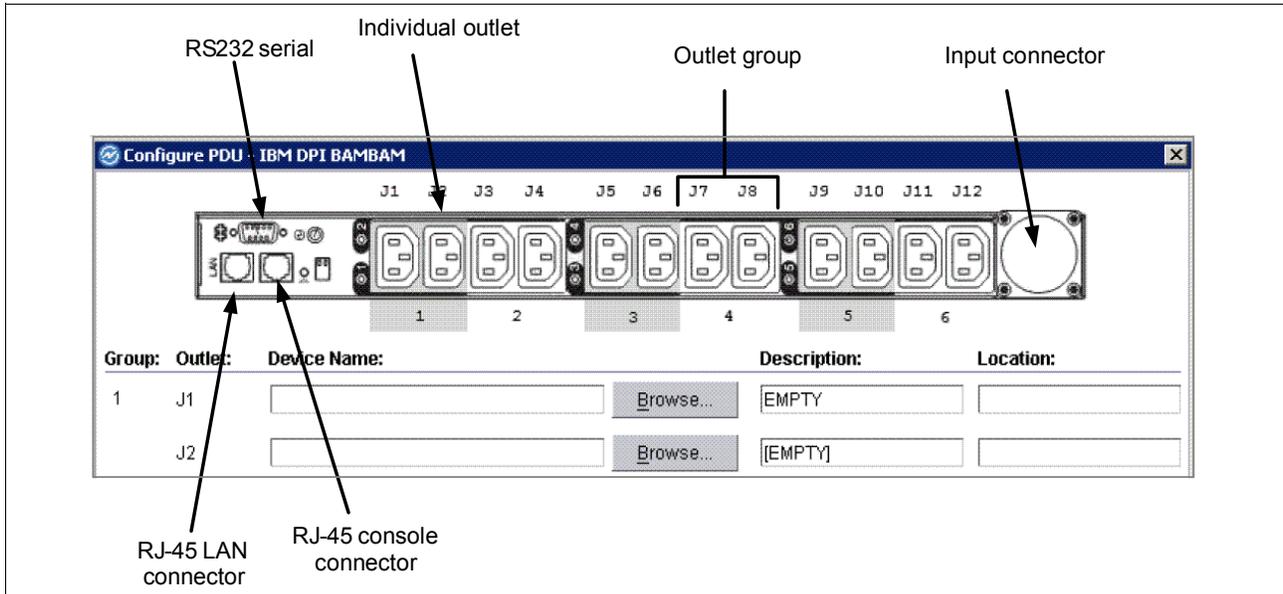
Figure 1-8   Power Distribution Unit plus example

In this panel, outlet names and outlet group names are shown. Each PDU+ node will contain either node group outlets or individual outlets.

Active Energy Manager runs as an extension to IBM Director 5.20.2. For more information about Active Energy Manager, refer to 2.2, "IBM EnergyScale technology" on page 62.

1.14 Rack-mounting rules

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

► The system 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 the system into the service position, it is essential that the rack manufacturer's safety instructions have been followed regarding rack stability.

1.15 Useful rack additions

This section highlights some solutions available for IBM System p rack-based systems.

IBM 7214 Model 1U2 SAS Storage Enclosure

The IBM System Storage 7214 Tape and Optical Storage Device Enclosure Model 1U2 (7214 Model 1U2) is designed to mount in one EIA unit of a standard 19 inch rack and can be configured with one or two tape drives, or either one or two Slim DVD-RAM or DVD-ROM drives in the right-side bay.

The 7214 Model 1U2 is designed to attach to specific models of the System p, including the 8203-E4A, through an external serial-attached SCSI (SAS) adapter. The SAS electronic interface offered on the IBM 7214 Model 1U2 is designed to provide faster transfer rates compared to SCSI transfer rates, and a reduction in space required for system-to-device cabling.

The 7214 Model 1U2 is intended to provide external tape or optical drive options that allows customers to transfer data with similar archival storage and retrieval technologies already installed in existing IT facilities.

Optional drive technologies available in the 7214 Model 1U2 offering are:

► DAT72 36 GB tape drive

► DVD-RAM — 9.4 GB Slim Optical Drive

► DVD-ROM — 9.4 GB Slim Optical Drive

► DAT72 36 GB data cartridges are read-write compatible with media used in previous DDS4 tape drives.

Most standard DVD disks can be used in the 7214 Model 1U2 DVD drive features. A control card sensor in the 7214 Model 1U2 can track drive function and notify the user of maintenance requirements.

The 7214 Model 1U2 offers IBM's Customer Replaceable Unit (CRU) maintenance service to help make installation or replacement of drives efficient. Other 7214 Model 1U2 components are also designed for CRU maintenance.

IBM 7212 Model 103 IBM System Storage Tape and DVD enclosure

The IBM 7212 System Storage Tape and DVD Enclosure Express Model 103 is designed to mount in one EIA Unit of a standard 19 inch rack, using an optional rack-mount hardware feature kit. The 7212 Express Model can be configured with one or two tape or DVD-RAM optical drives, and can also be configured for desktop installation. The base unit allows for a split-bus configuration for connection of each drive to individual external SCSI-LVD adapters on IBM System i and IBM System p servers. An optional single-bus configuration of the 7212-103 allows for connection of both drives to only one external SCSI-LVD adapter. The Express offering for the IBM 7212 Model 103 is intended to provide external tape or optical drive options that can be compatible with other archival storage and retrieval technologies already installed in existing IT facilities.

Optional drive technologies available in the 7212-103 Express offering are:

► 4 mm - DAT72 36 GB tape drive

► VXA - VXA-2 80 GB tape drive

► VXA-320 160 GB tape drive

- SLR (1/4 inch format) - SLR60 37.5 GB tape drive
- SLR100 50 GB Tape Drive
- LTO - HH-LTO2 200 GB tape drive
- DVD-RAM - 9.4 GB optical drive

Flat panel display options

The IBM 7316 Model TF3 is a rack-mountable flat panel console kit consisting of a 17 inch 337.9 mm x 270.3 mm flat panel color monitor, rack keyboard tray, IBM Travel Keyboard, support for IBM Keyboard/Video/Mouse (KVM) switches, and language support. The IBM 7316-TF3 Flat Panel Console Kit offers:

- Slim, sleek, lightweight monitor design that occupies only 1U (1.75 inches) in a 19 inch standard rack
- A 17 inch, flat screen TFT monitor with truly accurate images and virtually no distortion
- Ability to mount the IBM Travel Keyboard in the 7316-TF3 rack keyboard tray
- Support for IBM Keyboard/Video/Mouse (KVM) switches that provide control of as many as 128 servers, and support for both USB and PS/2 server-side keyboard and mouse connections

Note: We recommend that you have the 7316-TF3 installed between EIA 20 to 25 of the rack for ease of use.

1.15.1 OEM rack

The system 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 Site and Hardware Planning Information*, 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 in. + 0.03 in.), and the rail-mounting holes must be 465 mm + 0.8 mm (18.3 in. + 0.03 in.) 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 1-9 for a top view showing the specification dimensions.
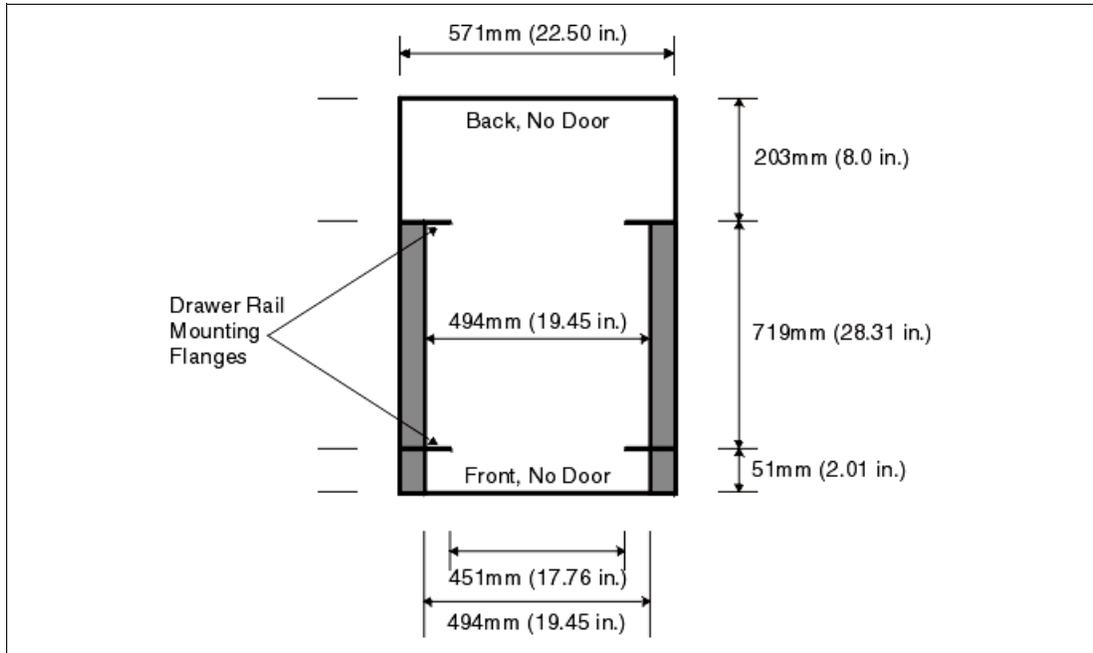
Figure 1-9 Top view of non-IBM rack specification dimensions

► The vertical distance between the mounting holes must consist of sets of three holes spaced (from bottom to top) 15.9 mm (0.625 in.), 15.9 mm (0.625 in.), and 12.67 mm (0.5 in.) on center, making each three-hole set of vertical hole spacing 44.45 mm (1.75 in.) apart on center. Rail-mounting holes must be 7.1 mm + 0.1 mm (0.28 in. + 0.004 in.) in diameter. See Figure 1-10 and Figure 1-11 for the top and bottom front specification dimensions.
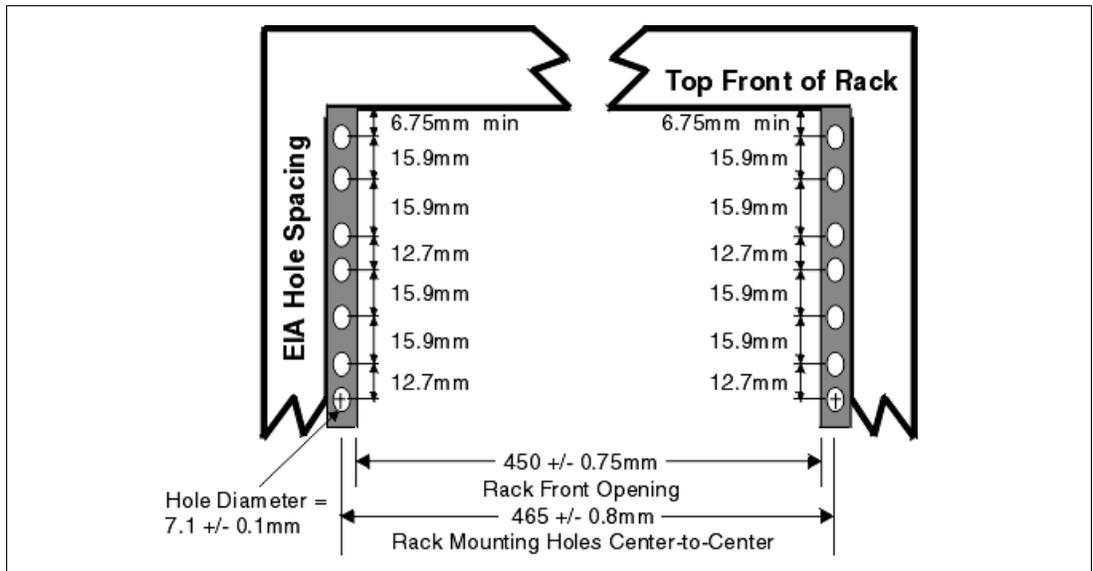
Figure 1-10 Rack specification dimensions, top front view
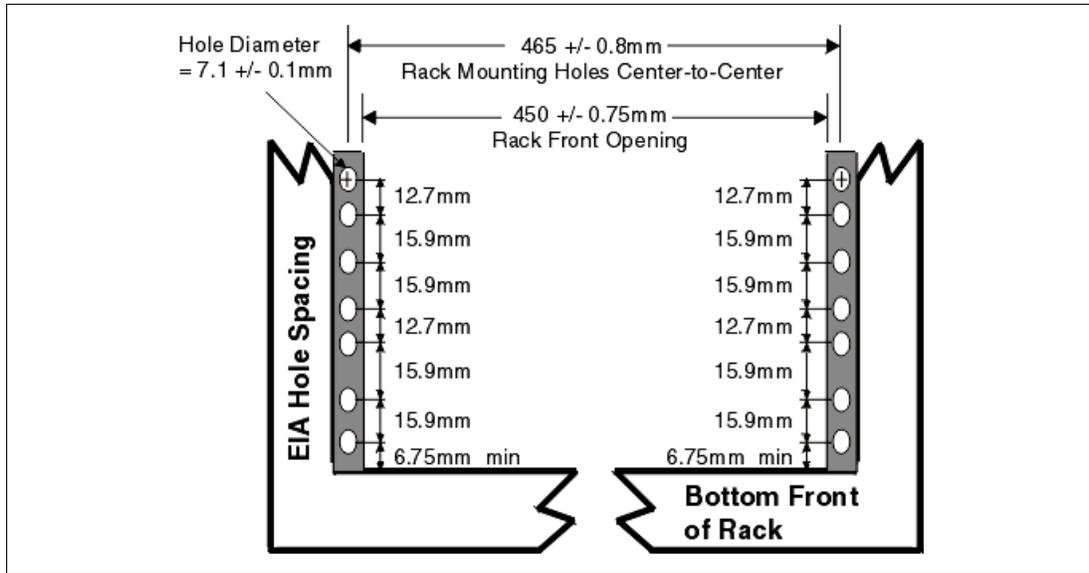
Figure 1-11   Rack specification dimensions, bottom front view

► 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 installed I/O hardware requirements.

1.16 Operating system support

The POWER6-based Power 520 requires the following levels of the supported AIX, IBM i, and Linux operating systems:

► For the 8203-E4A:

– AIX V5.3 with the 5300-06 Technology Level with Service Pack 7, or later

– AIX V5.3 with the 5300-07 Technology Level, or later

– AIX V6.1, or later

– IBM i 5.4 and IBM i 6.1, AIX, and Linux partitions
  For use with VIOS or IVM, IBM i 6.1 is required

– SUSE Linux Enterprise Server 10 (SLES 10) Service Pack 1 for Power, or later

– Red Hat Enterprise Linux V4.5 for Power, or later

– Red Hat Enterprise Linux V5.1 for Power, or later

- ► Supported IBM i console choices include:
  - – Operations console attached through an Ethernet port (LAN console) or WAN port ("ops console")
  - – Hardware Management Console (HMC)

A twinax console is not supported unless a Hardware Management Console (HMC) is present on the system. A 9944-100 Thin Console is not supported by IBM i on POWER6 systems.

IBM periodically releases fixes, group fixes, and cumulative fix packages for the IBM i operating system. These packages can be ordered on CD-ROM or downloaded from:

http://www-933.ibm.com/eserver/support/fixes/fixcentral

Select a product (hardware) family. For POWER6 select Power.

A sequence of selection fields appear as you select an entry for a previous field. Selection fields include selection of an operating system (for example AIX, IBM i, Linux) and other software categories that include microcode, firmware, and others. For most options you must select a release level.

The Fix Central Web site provides information about how to obtain the software using the media (for example, the CD-ROM).

You can also use the Fix Central Web site to search for and download individual operating system fixes, licensed program fixes, and additional information.

Part of the fix processing is Fix Central dialoguing with your IBM i or AIX operating system to identify fixes already installed, and whether additional fixes are required.

1.16.1 IBM AIX 5.3

If installing AIX 5L 5.3 on the 8203-E4A, either one of the following minimum requirements must be met:

- ► AIX 5L for POWER V5.3 with 5300-07 Technology Level or later
- ► IBM AIX V6.1 with T6100-01 Technology Level or later

IBM periodically releases maintenance packages (service packs or technology levels) for the AIX 5L operating system. These packages can be ordered on CD-ROM or downloaded from:

http://www-912.ibm.com/eserver/support/fixes/fixcentral/main/pseries/aix

Select the release level. The Fix Central Web site also provides information about how to obtain the software using the media (for example, the CD-ROM).

You can also get individual operating system fixes and information about obtaining AIX 5L service at this site. From AIX 5L V5.3 the Service Update Management Assistant is part of the base operating system. It helps the administrator to automate the task of checking and downloading operating system downloads. For more information about the `suma` command functionality, refer to:

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

AIX 5L is supported on the System p servers in partitions with dedicated processors (LPARs), and shared-processor partitions (micro-partitions). When combined with one of the PowerVM features, AIX 5L Version 5.3 or later can make use of all the existing and new virtualization features such as micro-partitions, virtual I/O, virtual LAN, PowerVM Live Partition Mobility, and so forth.

1.16.2 IBM AIX V6.1

IBM AIX V6.1 recently became available. It includes significant new capabilities for virtualization, security features, continuous availability features, and manageability. AIX V6.1 is the first generally available version of AIX V6.

AIX V6.1 features include support for:

► PowerVM AIX 6 Workload Partitions (WPAR) – software-based virtualization

► Live Application Mobility – with the IBM PowerVM AIX 6 Workload Partitions Manager for AIX (5765-WPM)

► 64-bit Kernel for higher scalability and performance

► Dynamic logical partitioning and Micro-Partitioning

► Multiple Shared-Processor Pools

► Trusted AIX – MultiLevel, compartmentalized security

► Integrated role-based access control

► Encrypting JFS2 file system

► Kernel exploitation of POWER6 Storage Keys for greater reliability

► Robust journaled file system and Logical Volume Manager (LVM) software including integrated file system snapshot

► Tools for managing the systems environment – System Management

– System Management Interface Tool (SMIT)
– IBM Systems Director Console for AIX

1.16.3 IBM i 5.4 (formerly IBM i5/OS V5R4)

IBM i 5.4 runs on POWER6, POWER5, and older technology System i models. Together with Licensed Internal Code (Machine Code) level 5.4.5, IBM i 5.4 is the minimum IBM i release level required on POWER6 technology systems. It contains a wide range of medium to small enhancements and new functions built on top of IBM i's integrated work management, performance management, database (DB2 for i5/OS), security, backup and recovery functions, and System i Navigator graphical interface to these functions.

5.4 enhancements include:

► Support for POWER6 processor technology models

► Support for large write cache disk controllers (IOAs)

► Expanded support for IOAs that do not require IOPs

► More flexible backup and recovery options and extended support in local remote journaling and cross-site mirroring and clustering

► Expanded DB2 and SQL functions and graphical management

► IBM Control Language (CL) extensions

► Initial release support for IBM Technology for Java 32-bit JVM™

► Support for IBM Express Runtime Web Environments for i5/OS, which contains a wide range of capabilities intended to assist new or inexperienced Web users to get up and running in a Web application serving environment

► Expanded handling of 5250 workstation applications running in a Web environment using the WebFacing and HATS components

- Licensed program enhancements, including Backup Recovery and Media Services, and application development enhancements, including RPG, COBOL, and C/C++.

1.16.4 IBM i 6.1

6.1 runs on POWER6, POWER5, and the last generation of System i 8xx models (800, 810, 825, 870, 890). As with previous releases, 6.1 builds on top of the IBM i integrated capabilities with enhancement primarily in the areas of:

- IBM i security, including greatly expanded data encryption/decryption and network intrusion detection.
- Support for the IBM PCI-X (#5749) and PCIe Fibre Channel (#5774) IOP-less adapters and a new performance improved code path for attached IBM System Storage DS8000 configurations.
- Expanded base save/restore, journaling, and clustering support.
- New IBM high availability products that take advantage of the expanded 6.1 save/restore, journaling, and clustering support.
- System i PowerHA for i (formerly known as High Availability Solutions Manager (HASM)) and IBM iCluster® for i.
- Logical partitioning extensions, including support for multiple shared processor pools and IBM i 6.1 as a client partition to another 6.1 server partition or a server IBM Virtual I/O Server partition. The VIOS partition can be on a POWER6 server or a POWER6 IBM Blade JS22 or JS12.
- Expanded DB2 and SQL functions, graphical management of the database, and generally improved performance.
- Integrated Web application server and Web Services server (for those getting started with Web services.
- Integrated browser-based IBM Systems Director Navigator for i5/OS that includes a new Investigate Performance Data graphically function.
- Initial release support for IBM Technology for Java 64-bit JVM.
- RPG, COBOL, and C/C++ enhancements, as well as new packaging of the application development tools WebSphere Development Studio and Rational Developer.

> **Note:** IBM i 6.1 takes maximum advantage of machine instructions used on POWER5 and POWER6 technology systems. Application programs compiled or created on 6.1 automatically take advantage of the restructured program object. For application programs created on prior releases being loaded (restored) onto a 6.1 partition, an object conversion (re-translation) must be performed prior to or at first execution of the program.
>
> IBM offers planning information and examples in the IBM Redbooks publication *IBM i5/OS Program Conversion: Getting Ready for i5/OS V6R1*, REDP-4293.

1.16.5 Linux systems summary

Linux is an open source operating system that runs on numerous platforms from embedded systems to mainframe computers. It provides a UNIX-like implementation across many computer architectures. This section discusses two brands of Linux to be run in partitions.

The supported versions of Linux systems include:

► Novell® SUSE Linux Enterprise Server 10 (SLES 10) Service Pack 1 for Power or later
► Red Hat Enterprise Linux V4.5 for Power or later
► Red Hat Enterprise Linux V5.1 for Power or later

The PowerVM features are supported in Version 2.6.9 and above of the Linux kernel. The commercially available latest distributions from Red Hat, Inc. (RHEL AS 5) and Novell SUSE Linux (SLES 10) support the IBM system p 64-bit architectures and are based on this 2.6 kernel series.

If you want to configure Linux partitions in virtualized POWER systems, take into account the following considerations:

► Not all devices and features that are supported by the AIX operating system are supported in logical partitions running the Linux operating system.

► Linux operating system licenses are ordered separately from the hardware. You can acquire Linux operating system licenses from IBM, to be included with their Power 520, or from other Linux distributors.

Note: Linux for POWER systems clients can obtain a one-year or three-year renewal subscription for Red Hat Enterprise Linux (RHEL) from IBM. Previously, POWER systems clients who purchased one-year or three-year RHEL subscription from IBM were required to go directly to Red Hat, Inc. when their original one-year or three-year subscription expired. From 1st May 2009, IBM clients can obtain a one-year or three-year renewal subscription from IBM. This enhancement is only available for Red Hat Enterprise Linux for POWER.

Also, clients with an installed Power Systems server who did not order SUSE Linux Enterprise Server for POWER (SLES) at the time of the Power Systems hardware purchase can order SLES for that system from IBM. This enhancement is only available for SUSE Linux Enterprise Server for Power from 1st May 2009. Previously, IBM clients could only obtain a SLES operating system subscription from IBM when it was purchased with IBM hardware or with a processor upgrade for installed hardware.

IBM also offers the option to have SUSE Linux Enterprise Server 10 for POWER preloaded on select Power Systems servers. Save time and leverage IBM expertise by allowing IBM to install Linux in the factory prior to shipping the server. This option is available for IBM Power blades through Power 570 servers.

The Program Numbers for Subscription and Support for RHEL and SLES are:

```
Red Hat Enterprise Linux AP for Power : 5639-RHL
SUSE Linux Enterprise Server 10 for Power : 5639-SLP
```

IBM only supports the Linux systems of clients with a Support Line contract covering Linux. Otherwise, contact the Linux distributor for support.

More information

For information about the features and external devices supported by Linux refer to:

http://www-03.ibm.com/systems/p/os/linux/index.html

For information about SUSE Linux Enterprise Server 10, refer to:

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

For information about Red Hat Enterprise Linux Advanced Server 5, refer to:

http://www.redhat.com/rhel/features

1.16.6 Supported virtualization features with Linux

SLES 10, RHEL AS 4.5, and RHEL AS 5 support the following virtualization features:

► Virtual SCSI, including for the boot device
► Shared-processor partitions and virtual processors, capped and uncapped
► Dedicated-processor partitions
► Dynamic reconfiguration of processors
► Virtual Ethernet, including connections through the Shared Ethernet Adapter in the Virtual I/O Server to a physical Ethernet connection
► Simultaneous multithreading

SLES 10, RHEL AS 4.5, and RHEL AS 5 do *not* support the following features:

► Dynamic reconfiguration of memory
► Dynamic reconfiguration of I/O slot

2

Architecture and technical overview

This chapter discusses the overall system architecture represented by Figure 2-1 on page 58, with its major components described in the following sections. The bandwidths that are provided throughout are theoretical maximums used for reference.

The speeds shown are at an individual component level. Multiple components and application implementation are key to achieving the best performance.

You should always do performance sizing at the application workload environment level and evaluate performance using real-world performance measurements using production workloads.
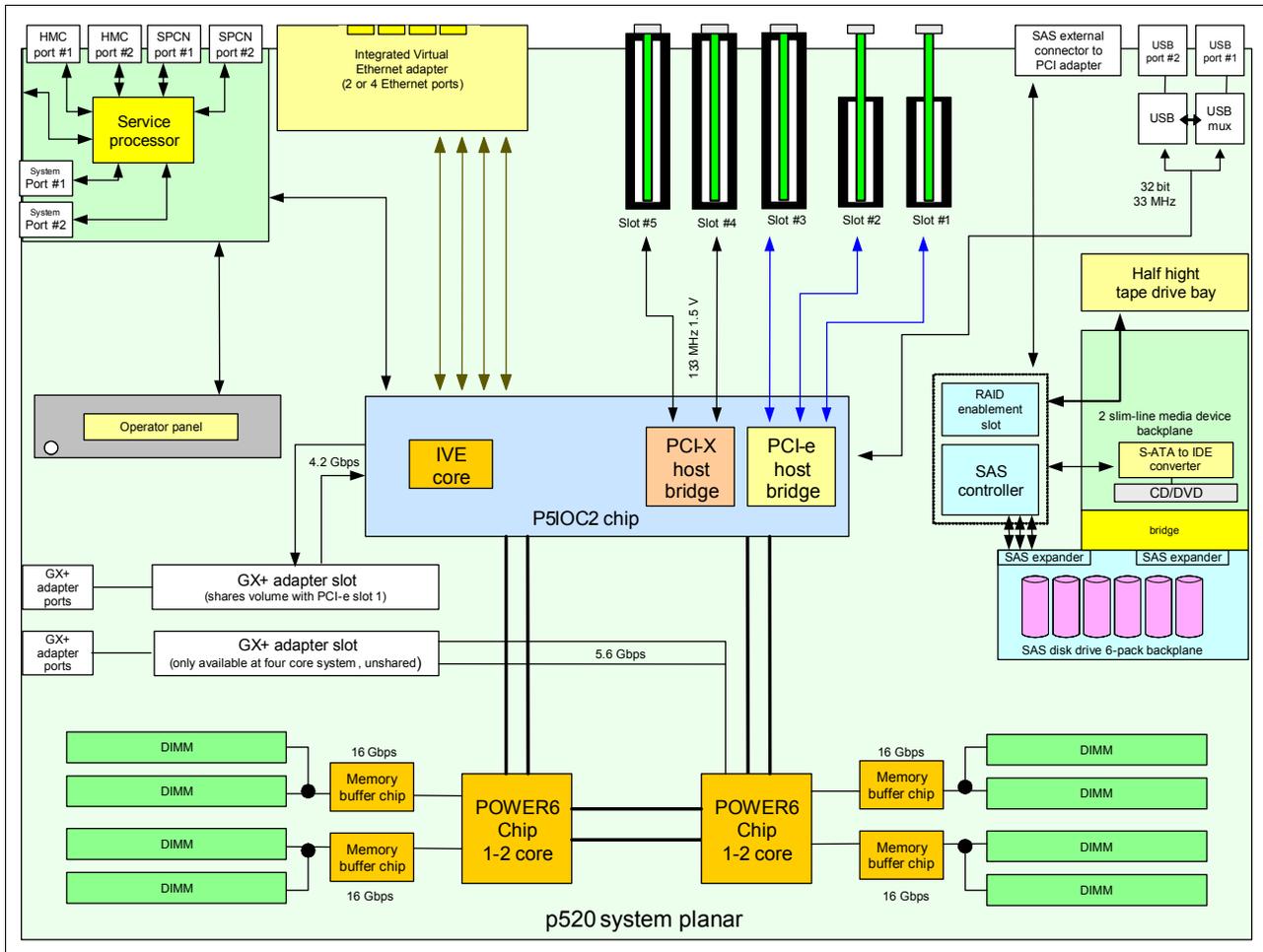
Figure 2-1   Power 520 logic data flow (8310 backplane)

2.1 The POWER6 processor

The POWER6 processor capitalizes on all the enhancements brought by the POWER5 processor.

Two of the enhancements of the POWER6 processor are the ability to do processor instruction retry and alternate processor recovery. This significantly reduces exposure to both hard (logic) and soft (transient) errors in the processor core.

► **Processor instruction retry**
Soft failures in the processor core are transient errors. When an error is encountered in the core, the POWER6 processor will first automatically retry the instruction. If the source of the error was truly transient, the instruction will succeed and the system will continue as before. On predecessor IBM systems, this error would have caused a checkstop.

► **Alternate processor retry**
Hard failures are more difficult, being true logical errors that will be replicated each time the instruction is repeated. Retrying the instruction will not help in this situation because the instruction will continue to fail. Systems with POWER6 processors introduce the ability to extract the failing instruction from the faulty core and retry it elsewhere in the system, after which the failing core is dynamically deconfigured and called out for replacement.

The entire process is transparent to the partition owning the failing instruction. Systems with POWER6 processors are designed to avoid what would have been a full system outage.

▶ **POWER6 single processor checkstopping**
Another major advancement in POWER6 processors is single processor checkstopping. A processor checkstop would result in a system checkstop. A feature in Power 520 is the ability to contain most processor checkstops to the partition that was using the processor at the time. This significantly reduces the probability of any one processor affecting total system availability.

▶ **POWER6 cache availability**
In the event that an uncorrectable error occurs in L2 or L3 cache (not available on the Power 520), the system will be able to dynamically remove the offending line of cache without requiring a reboot. In addition, POWER6 utilizes an L1/L2 cache design and a write-through cache policy on all levels, helping to ensure that data is written to main memory as soon as possible.

Figure 2-2 shows a high level view of the POWER6 processor. L1 Data and L1 Instruction caches are within the POWER6 core.
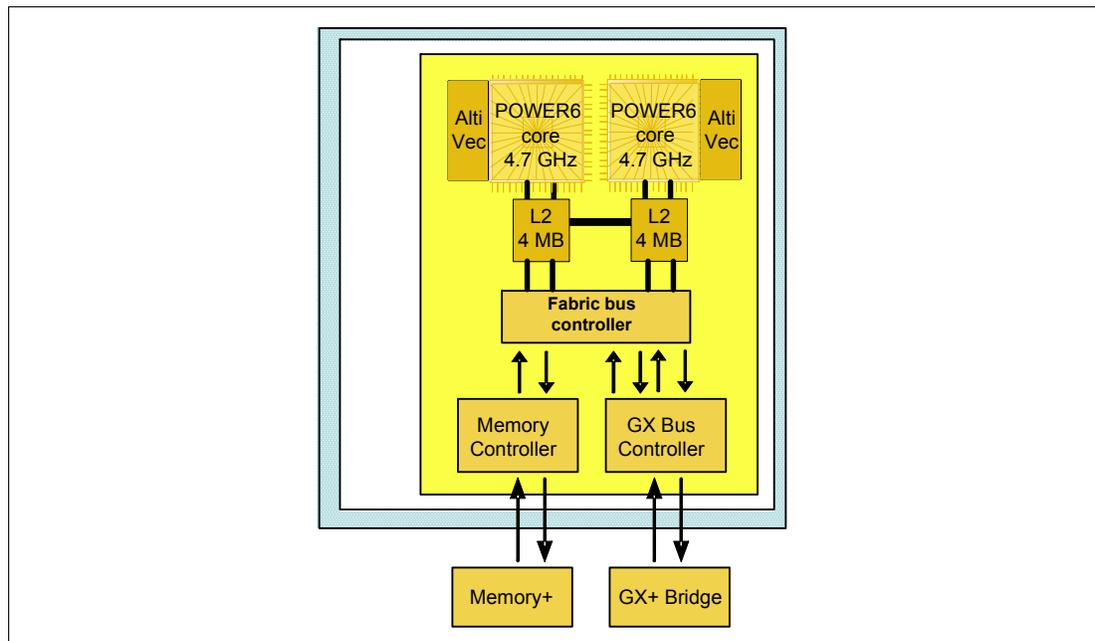
Figure 2-2   POWER6 processor

The CMOS 11S0 lithography technology in the POWER6 processor uses a 65 nm fabrication process, which enables:

▶ Performance gains through faster clock rates – from 4.2 GHz to 4.7 GHz.
▶ Physical size of 341 mm

The POWER6 processor consumes less power and requires less cooling. Thus, you can use the POWER6 processor in servers where previously you could only use lower frequency chips due to cooling restrictions.

The 64-bit implementation of the POWER6 design provides the following additional enhancements:

- ► Compatibility of 64-bit architecture
  - – Binary compatibility for all POWER and PowerPC® application code level
  - – Support for partition migration
  - – Support Big and little Endian
  - – Support for four page sizes: 4 KB, 64 KB, 16 MB, and 16 GB
- ► High frequency optimization
  - – Designed to operate at maximum speed of 5 GHz
- ► Superscalar core organization
  - – Simultaneous Multithreading: two threads
- ► In-order dispatch of five operations (through a single thread) or seven operations (using Simultaneous Multithreading) to nine execution units:
  - – Two load or store operations
  - – Two fixed-point register-register operations
  - – Two floating-point operations
  - – One branch operation

The POWER6 processor implements the 64-bit IBM Power Architecture® technology. Each POWER6 chip incorporates two ultrahigh dual-threaded Simultaneous Multithreading processor cores, a private 4 MB level 2 cache (L2) for each processor, integrated memory controller and data interconnect switch and support logic for dynamic power management, dynamic configuration and recovery, and system monitoring.

2.1.1 Decimal floating point

This section describes the behavior of the POWER6 hardware decimal floating-point processor, the supported data types, formats, and classes, and the use of registers. Prior to POWER6 technology most technologies implemented decimal floating point in software.

The decimal floating-point (DFP) processor shares the 32 floating-point registers (FPRs) and the floating-point status and control register (FPSCR) with the binary floating-point (BFP) processor. However, the interpretation of data formats in the FPRs, and the meaning of some control and status bits in the FPSCR are different between the BFP and DFP processors.

The DFP processor supports three DFP data formats:

- ► DFP32 (single precision): 4 bytes, 7 digits precision, -95/+96 exponent
- ► DFP64 (double precision): 8 bytes, 16 digits precision, -383/+384 exponent
- ► DFP128 (quad precision): 16 bytes, 34 digits precision, -6143/+6144 exponent

Most operations are performed on the DFP64 or DFP128 format directly. Support for DFP32 is limited to conversion to and from DFP64. For some operations, the DFP processor also supports operands in other data types, including signed or unsigned binary fixed-point data, and signed or unsigned decimal data.

DFP instructions are provided to perform arithmetic, compare, test, quantum-adjustment, conversion, and format operations on operands held in FPRs or FPR pairs.

- ► Arithmetic instructions: These instructions perform addition, subtraction, multiplication, and division operations.

- ► Compare instructions: These instructions perform a comparison operation on the numerical value of two DFP operands.

- ► Test instructions: These instructions test the data class, the data group, the exponent, or the number of significant digits of a DFP operand.

- ► Quantum-adjustment instructions: These instructions convert a DFP number to a result in the form that has the designated exponent, which may be explicitly or implicitly specified.

- ► Conversion instructions: These instructions perform conversion between different data formats or data types.

- ► Format instructions: These instructions facilitate composing or decomposing a DFP operand.

Enabling applications running on POWER6 systems to take advantage of the hardware decimal floating point support depends on the programming language release level used by the application and the operating system in which the application is running.

The following list includes examples:

- ► Java applications: Applications running IBM Technology for Java 6.0 32-bit and 64-bit JVM automatically take advantage of the hardware assist during the initial Just in Time (JIT) processing. Applications running under IBM i require release level 6.1; Java 5.0 does not use DFP.

    C and C++ applications: For the C and C++ compilers running under AIX and Linux systems, as of v9.0, there is DFP support through the POWER6 hardware instructions. Software emulation is supported on all other POWER architectures.

    Running under IBM i 6.1, support for DFP has been added to the IBM i 6.1 ILE C compiler. If a C program that uses DFP data is compiled on POWER 6 hardware, hardware DFP instructions will be generated; otherwise, software emulation will be used.

    IBM i support for DFP in the ILE C++ compiler is planned for a future release.

    For your information, C and C++ on z/OS®, as of V1R9, use hardware DFP support where the runtime code detects hardware analogous to POWER 6.

- ► IBM i ILE RPG and COBOL: These languages do not use decimal floating point. The normal zoned decimal or packed decimal instructions receive "normal" performance gains merely by running under IBM i 6.1 on POWER6.

    IBM i 6.1 supports decimal floating point data, for example, in DB2 for i5/OS tables. If the RPG or COBOL compiler encounters a decimal float variable in an externally-described file or data structure, it will ignore the variable and issue an identifying information message.

- ► Some applications, such as those available from SAP, that run on POWER6-based systems, might provide specific ways to take advantage of decimal floating point.

    For example, the SAP NetWeaver 7.10 ABAP™ kernel introduces a new SAP ABAP data type called DECFLOAT to enable more accurate and consistent results from decimal floating point computations. The DFP support by SAP NetWeaver leverages the built-in DFP feature of POWER6 processors. This allows for simplified ABAP coding while increasing numeric accuracy and adding the potential for significant performance improvements.

2.1.2 AltiVec and Single Instruction, Multiple Data

IBM Semiconductor's advanced Single Instruction, Multiple Data (SIMD) technology based on the AltiVec instruction set is designed to enable exceptional general purpose processing

power for high performance POWER processors. This technology is engineered to support high bandwidth data processing and algorithmic-intensive computations, all in a single chip solution.

With its computing power, AltiVec technology also enables high performance POWER processors to address markets and applications in which performance must be balanced with power consumption, system cost, and peripheral integration.

The AltiVec technology is a well known environment for software developers who want to add efficiency and speed to their applications. A 128-bit vector execution unit was added to the architecture. This engine operates concurrently with the existing integer and floating-point units and enables highly parallel operations – up to 16 operations in a single clock cycle. By leveraging AltiVec technology, developers can optimize applications to deliver acceleration in performance driven, high bandwidth computing.

The AltiVec technology is not comparable to the IBM POWER6 processor implementation because it uses the Simultaneous Multithreading functionality.

2.2 IBM EnergyScale technology

IBM EnergyScale technology is featured on the IBM POWER6 processor-based systems. It provides functions to help the user understand and control IBM server power and cooling usage.

This section describes the IBM EnergyScale features and hardware and software requirements.

Power trending   EnergyScale provides continuous power usage data collection (monitoring). This enables administrators to predict power consumption across their infrastructure and to react to business and processing needs. For example, an administrator could adjust server consumption to reduce electrical costs. To collect power data for the Power 520 you do not need any additional hardware because the information is collected internally.

Energy saver mode Energy (formerly Power) saver mode reduces the voltage and frequency by a fixed percentage. This percentage is predetermined to be within a safe operating limit and is not user configurable. Under current implementation this is a 14% frequency drop. When CPU utilization is low, Energy Saver Mode has no impact on performance. Energy Saver Mode can reduce the processor usage up to 30%. Energy Saver Mode is not supported during boot or reboot although it is a persistent condition that will be sustained after the boot when the system starts executing instructions. Energy Saver is only supported with 4.0 GHz processors and faster.

Energy capping   Energy (formerly Power) capping enforces a user-specified limit on power usage. Energy Capping is not a power saving mechanism. It enforces power caps by actually throttling the processors in the system, degrading performance significantly. The idea of a power cap is to set something that should never be reached but frees up *margined power* in the data center. The margined power is the amount of extra power that is allocated to a server during its installation in a datacenter. It is based on the server environmental specifications that usually are never reached. This is because server specifications are always based on maximum configurations and worst case scenarios.

Processor core nap The IBM POWER6 processor uses a low-power mode called Nap that stops processor execution when there is no work to do on that processor core (both threads are idle). Nap mode allows the hardware to clock off most of the circuits inside the processor core. Reducing active power consumption by turning off the clocks allows the temperature to fall, which further reduces leakage (static) power of the circuits, causing a cumulative effect. Unlicensed cores are kept in core Nap until they are licensed, and return to core Nap whenever they are unlicensed again.

EnergyScale for I/O IBM POWER6 processor-based systems automatically power off pluggable, PCI adapter slots that are empty or not being used, which saves approximately 14 watts per slot. System firmware automatically scans all pluggable PCI slots at regular intervals, looking for those that meet the criteria for being not in use and powering them off. This support is available for all POWER6 processor-based servers, and the expansion units that they support. Note that it applies to hot pluggable PCI slots only.

2.2.1 Thermal power management device (TPMD)

The implementation of performance-aware power and thermal management for POWER6 processor based systems is called the EnergyScale architecture. The EnergyScale architecture meets a number of basic requirements for system level power. The Power 520 implementation uses an optional plug-in card containing a microcontroller called the thermal power management device (TPMD).

The TPMD card is part of the energy management of performance and thermal proposal, which dynamically optimizes the processor performance depending on processor power and system workload.

Figure 2-3 depicts the TPMD card within the system enclosure.
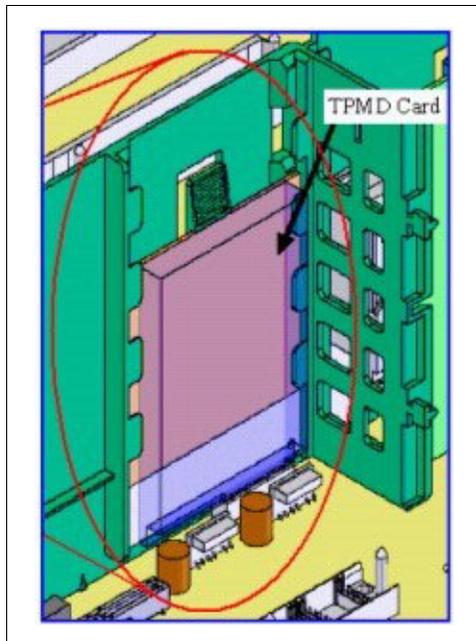
Figure 2-3   Thermal power management device (TPMD) card

The EnergyScale design supports a number of power and thermal management policies, including:

► **Benchmark:** Benchmark maximizes the single-threaded performance of the system by putting one core of each processor into a low-power state, like the POWER6 nap mode.

► **Maximum performance**: The EnergyScale implementation regulates the system in such a way that it achieves the maximum performance possible without violating the power or thermal limits of the system.

► **Power cap:** The system will be set to a previously defined power usage limit.

► **Maximum power savings**: EnergyScale attempts to save as much power as possible for a given workload.

► **Optimal power/performance**: In this mode the EnergyScale implementation changes the system to the most optimal power/performance settings on the basis of workload characteristics and the power and thermal environment.

> **Note:** A TPMD card is standard for an initial order. The system will not boot without an installed TPMD card.

EnergyScale architecture is implemented by firmware running on the TPMD and the SP of the system and the IBM Systems Director Active Energy Manager management software, which is a plug-in to the IBM Director software.

2.2.2 Hardware and software requirements

This sections summarizes the supported systems and software user interfaces for EnergyScale functions.

Table 2-1   EnergyScale systems support

Server	Power trending	Energy saver mode	Energy (power) capping	Processor Nap	I/O
BladeCenter JS22 Express 7998-61X	Y	Y	Y	Y	N
8203-E4A	Y	Y	Y	Y	Y
8204-E8A	Y	Y	Y	Y	Y
9117-MMA (< 4.0 GHz)	Y - via iPDU[a]	N	N	Y	Y
9117-MMA (>= 4.0 GHz)	Y - via iPDU[a]	Y[b]	N	Y	Y

a. An iPDU is required for this support. The feature code 5889 is for a base iPDU in a rack while the feature code 7109 is for additional iPDUs for the same rack. Supported racks are 7014-B42, 7014-S25/0555, 7014-T00/0551, and 7014-T42/0553.
b. Only supported if GX Dual Port RIO-2 Attach (#1800) is not present.

The primary user interface for EnergyScale features on a POWER6 based system is IBM Systems Director Active Energy Manager running as an extension to IBM Director 5.20.2 or later.

Table 2-2 on page 65 shows the ASMI, HMC, and Active Energy Manager interface support.

Table 2-2   EnergyScale function's software interfaces

EnergyScale functions	ASMI	HMC	Active Energy Manager
Power Trending	N	N	Y
Energy (Power) Saver Mode	Y	Y	Y
Schedule Energy Saver Mode Operation	N	Y	Y
Energy (POwer) Capping	N	N	Y
Schedule Energy Capping Operation	N	N	Y

To learn more about the capabilities of the IBM EnergyScale for POWER6 based systems and Active Energy Manager, refer to:

► Active Energy Manager Web site (as an extension to IBM Director):

  http://www–03.ibm.com/systems/management/director/extensions/actengmrg.htm]

► IBM Whitepaper, IBM EnergyScale for POWER6 Processor-Based Systems, April 2008, or later, which can be linked to from:

  http://www–03.ibm.com/systems/p/hardware/whitepapers/energyscale.html

► IBM Redbooks publication *Going Green with IBM Systems Director Active Energy Manager*, REDP-4361, at:

  http://www.ibm.com/redbooks

2.3 Processor cards

In the Power 520 the POWER6 processors and memory DIMMs are packaged directly on the system planar. The Power 520 1-core, 2-core, and 4-core systems use different POWER6 processor modules.

Note: Because the POWER6 processor modules are soldered directly to the system planar, you must take special care in sizing and selecting the ideal CPU configuration at initial order. You can take advantage of CoD for un activated processor cores, but you cannot upgrade out of a maximum processor Power 520 into a Power 550 or higher model.

Similarly, do appropriate sizing for maximum memory at initial order. Depending upon the number of processor cards (modules), ordering additional memory DIMMs later might cause you to have to discard the initial memory DIMMs when upgrading the amount of memory.

2.3.1 The Power 520 single-core module

The 1-core Power 520 system planar contains a single-core module and the local memory storage subsystem for that single-core. The 1-core 4.2 GHz system planar contains a single-core module and the local memory storage subsystem for it. L3 Cache is not available in this configuration. Figure 2-4 on page 66 shows the layout of a 4.2 GHz Power 520 and associated memory.
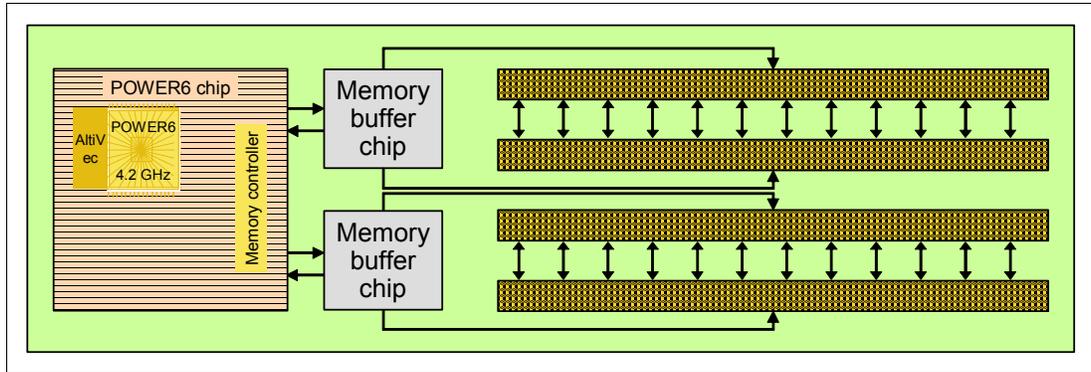
Figure 2-4   Power 520 memory socket layout view

This figure applies to both the 1-core 8203-E4A and the 9407-M15.

2.3.2 The Power 520 dual-core module

The 2-core Power 520 system planar contains a dual-core processor module and the local memory storage subsystem. The 4-core Power 520 (8203-E4A) system planar contains two dual-core processor modules, each with its physically plugged memory. All memory is accessible from all active processors.

Figure 2-5 shows a layout view of a Power 520 dual-core processor and associated memory.
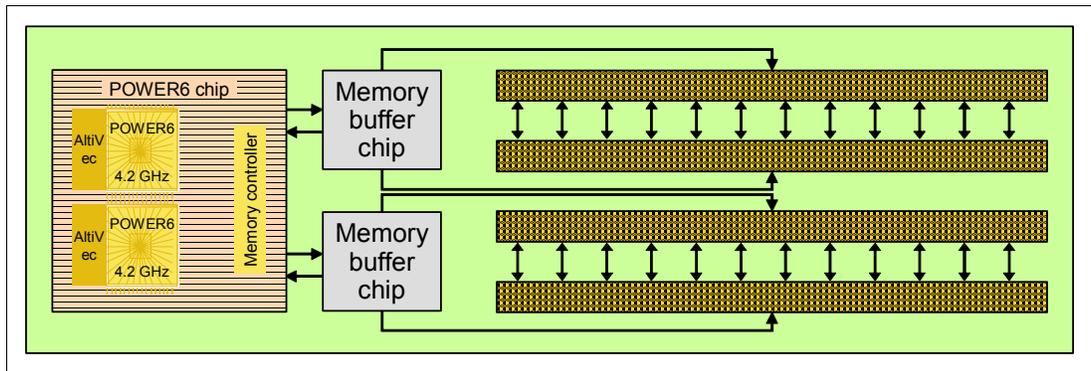
Figure 2-5   Power 520 4.2 GHz memory socket layout view

The storage structure for the POWER6 processor is a distributed memory architecture that provides high-memory bandwidth, although each processor can address all memory and sees a single shared memory resource. The processors are interfaced to four memory slots, controlled by two memory controller chips, which are located in close physical proximity to the processor modules.

I/O connects to the Power 520 processor module using the GX+ bus. The 2-core processor module configuration provides a single GX+ bus or PCIe x8 functionality using slot 1. Only a 4-core processor configuration provides a second, unshared GX+ connection.

The GX+ bus provides an interface to I/O devices through the RIO-2 connections.

This applies to the 2-core 8203-E4A.

The addition of a POWER6+ processor adds 32 MB to each dual-core processor, therefore a four-core processor card will have 64 MB of L3.

Note: The GX+ slot is not available on a 1-core configuration.

2.3.3 Processor clock rate

The Power 520 system features 1-core, 2-core, and 4-core configurations with the POWER6 processor running at 4.2 GHz or POWER6+ running at 4.7 GHz.

To verify the processor characteristics on a system running at 4.2 GHz, use one of the following commands:

▶ `lsattr -El procX`

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:

```
frequency    4208000000         Processor Speed        False
smt_enabled  true               Processor SMT enabled  False
smt_threads  2                  Processor SMT threads  False
state        enable             Processor state        False
type         powerPC_POWER6     Processor type         False
```

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

▶ `pmcycles -m`

The `pmcycles` command (available with AIX) uses the performance monitor cycle counter and the processor real-time clock to measure the actual processor clock speed in MHz. The following output is from a 2-core Power 520 system running at 4.2 GHz with simultaneous multithreading enabled:

```
Cpu 0 runs at 4208 MHz
Cpu 1 runs at 4208 MHz
Cpu 2 runs at 4208 MHz
Cpu 3 runs at 4208 MHz
```

Note: The `pmcycles` command is part of the bos.pmapi fileset. Use the `lslpp -l bos.pmapi` command to determine if it is installed on your system.

2.4 Memory subsystem

In the Power 520, the memory controller is internal to the POWER6 processor and it is interfaced to memory buffer chips that are located on the system board. The memory buffer chip in the POWER6 processor-based server is the next generation Synchronous Memory Interface ASIC that connects memory controllers to DDR-II memory DIMMs.

Any POWER6 chip interfaces two memory buffer chips, so the fully configured Power 520 has two POWER6 chips (for a total of four POWER6 processor cores) with the related memory buffer chips (see Figure 2-6 on page 68). Each channel from POWER6 chip to memory buffer chip allows 2 Bytes of Read data and 1 Byte of Write data.
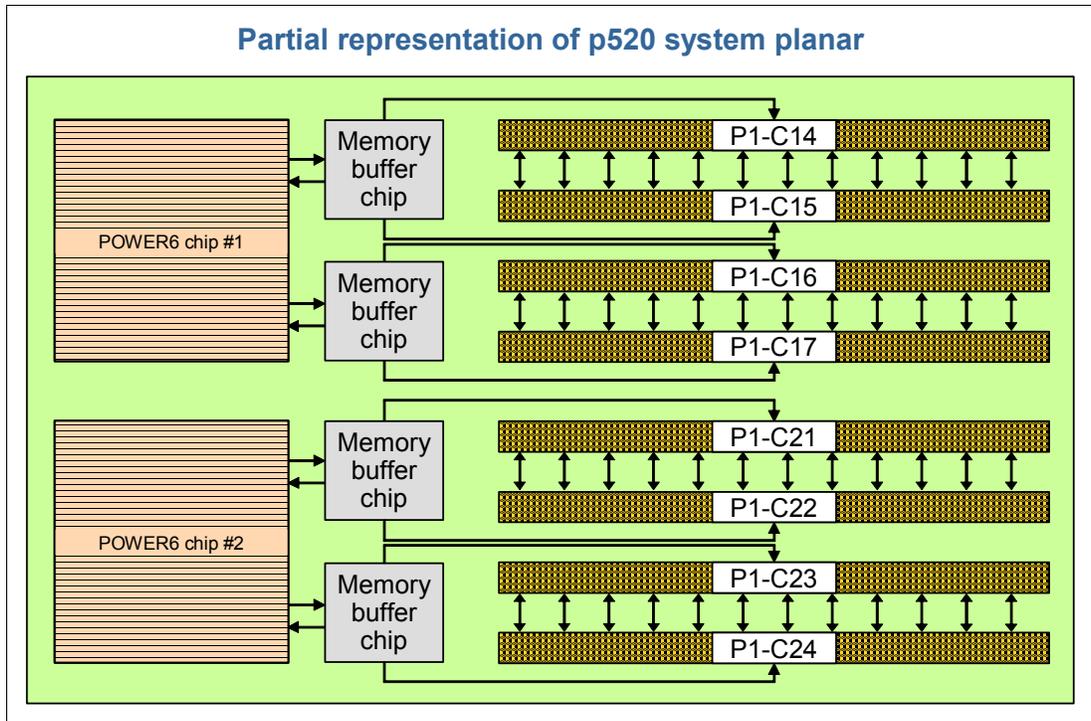
Figure 2-6   Power 520 memory subsystem and memory DIMM slots location

The supported memory module is the Industry Standard RDIMM, 240-pin, type PC2-3200 (DDR2 400 MBps) or PC2-5300 (DDR2 667 MBps) with enhanced memory RAS features including ECC, bit steering, chipkill, and resilience.

2.4.1 Memory placement rules

In the Power 520, the memory DIMMs are supported in pairs or quads. The minimum memory requirement is 2 GB.

Placement rules are dependant on the number of POWER6 chips configured.

Memory placement rules, one or two processor-core configuration

When configured with one POWER6 chip, with one or two processor cores enabled, the memory DIMMs can be installed in pairs or quad.

When DIMMs are installed in pairs, the priority to install the pairs is the following:

► The first DIMM pair should be installed in P1-C14 and P1-C16 slots

► The second DIMM pair should be installed in P1-C15 and P1-C17 slots

One POWER6 chip installed allows a maximum of one memory feature; the four memory modules from that feature should be installed in P1-C14, P1-C16, P1-C15 and P1-C17 slots.

All the memory DIMMs in a pair or in a quad must be of the same size and speed.

Memory placement rules, four processor-core configuration

When two POWER6 chips are available, with four processor-cores configured, all the connections to the memory slots are enabled. Then the memory can be installed in pairs or quads.

When DIMM are installed in pairs, the priority to install the pairs is as follow:

► The first pair should be installed in P1-C14 and P1-C16 that are connected to the first POWER6 chip through the related memory buffer chip modules

► The second pair should be installed in P1-C21 and P1-C23 that are connected to the second POWER6 chip through the related memory buffer chip modules

► The third pair should be installed in P1-C15 and P1-C17 (connected to the first POWER6 chip)

► The fourth pair should be installed in P1-C22 and P1-C24 (connected to the second POWER6 chip)

Two POWER6 chips installed allows a minimum of one quad and a maximum of two quads. The priority of the quads is as follows:

► The first quad requires two memory modules installed in P1-C14 and P1-C16 (connected to first POWER6 chip), then the other two memory modules installed in P1-C21 and P1-C23 (connected to the second POWER6 chip)

► The second quad requires two memory modules installed in P1-C15 and P1-C17 (related to the first POWER6 chip), then the other two modules in P1-C22 and P1-C24 (related to the second POWER6 chip)

All the memory DIMMs in a pair or in a quad must be of the same size and speed. The two different quads are separated and independent and can have different size and different speed memory. This is not recommended because, in operation, all memory accesses will tend to function at the slowest speed memory.

2.4.2 OEM memory

OEM memory is not supported or certified for use in IBM System p servers. If the Power 520 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 that is printed with a barcode and an alphanumeric string, as illustrated in Figure 2-7.
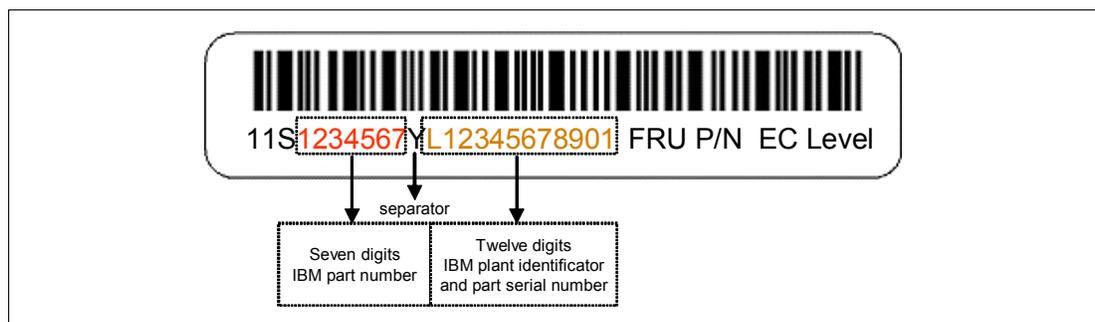


Figure 2-7   IBM memory certification label

2.4.3 Memory throughput

When ordering the IBM Power 520 MTM 8203-E4A, keep in mind the following memory throughput rules:

► The memory subsystem throughput is based on the speed of the memory.

- On the processor, there are two memory channels, each with single 2 byte read and 1 byte write. Memory channels of POWER6 memory controller are connected to memory buffers.

- The processor chip has two POWER6 processors.

- The DDR2 bus allows double reads or writes per clock cycle. If a 667 MHz memory feature is selected, the throughput is (2 x 2 x 2 x 2 x 667) + (2 x 1 x 2 x 2 x 667) or 16008 MBps or 16 GBps. These values are maximum theoretical throughputs for comparison purposes only.

Table 2-3 provides the theoretical throughput values for 4.2 GHz and 4.7 GHz processors with a 667 Mhz memory configuration.

Table 2-3   Theoretical throughput values.

Memory	Bandwidth for 4.2 GHz	Bandwidth for 4.7 GHz
L1 (Data)	67.2 GBps	75.2 GBps
L2 / Chip	168 GBps	188 GBps
Memory / Chip 4-cores	16 GBps 32 GBps	16 GBps 32 GBps

2.5 I/O capabilities

Each POWER6 processor configuration, except the single core configuration, provides a GX+ bus, which is used to connect to an I/O subsystem or Fabric Interface card. GX+ is a technology improvement over previous GX bus implementation. The processor module populating the first processor location is connected to the GX+ multifunctional host bridge chip, which provides the following major interfaces:

- One GX+ pass through bus: GX+ passthru elastic interface runs at one half the frequency of the primary. It allows other GX+ bus hubs to be connected into the system.

- Two 64-bit PCI-X 2.0 buses, one 64-bit PCI-X 1.0 bus, and one 32-bit PCI-X 1.0 bus

- Four 8x PCI Express links

- Two 10 Gbps Ethernet ports: Each port is individually configurable to function as two 1 Gbps port (depending on IVE feature selected)

The GX+ multifunctional host bridge provides a dedicated GX+ bus routed to the first GX+ slot through GX+ passthru bus. The second GX+ slot is available only on a 4-core configuration.

Optional Dual port RIO-2 I/O Hub (#1800) and Dual port 12x Channel Attach (#1802) that is installed in GX+ slot are used for external DASD and IO drawer and tower expansion in GX+ slot (Table 2-4). All GX+ cards are hot pluggable.

Table 2-4   GX+ adapters

Feature Code	Description	Slot	Size	Max
5609	GX Dual-port 12X Channel Attach	GX	GX slot	2
5614	Dual-port RIO-2 I/O Hub	GX	GX slot	2
5616	Dual-port 12X Channel Attach	GX	GX slot	2

When one GX+ slot is used, it overlays PCIe slot 1.

> **Note:** One GX+ is not available on a 1-core configuration. The GX+ adapter, if used, uses the PCIe x8 slot 1.
>
> The GX+ multifunctional host bridge chip is also commonly referred to in this publication as a PCI-X host bridge component for PCI-X interfaces or PCI-e host bridge component for PCIe interfaces. Readers with a System i background would refer to the PCI-X host bridge component as a Multi-Adapter Bridge.

Table 2-5 lists the I/O bus bandwidth of the 4.2 GHz and 4.7 GHz processors configurations.

Table 2-5   I/O bandwidth

I/O	Bandwidth for 4.2 GHz	Bandwidth for 4.7 GHz
Total	14 GBps	28.2 GBps
Primary GX Bus	8.4 GBps Bus with 4.2 GBps Passthru Bus	9.4 GBps Bus with 4.2 GBps Passthru Bus
GX Bus Slot 1	4.2 GBps (Passthru Bus)	4.7 GBps (Passthru Bus)
GX Bus Slot 2	5.6 GBps (4-core system)	18.8 GBps (4-core system)

2.6 Internal I/O subsystem

The internal system unit I/O subsystem resides on the system planar, which supports a mixture of both PCIe and PCI-X slots. All PCIe and PCI-X slots are hot pluggable and have Enhanced Error Handling (EEH). In the unlikely event of a problem, EEH-enabled adapters respond to a special data packet generated from the affected PCIe or 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.

Table 2-6 shows the Power 520 PCI slot configuration.

Table 2-6   Slot configuration of a Power 520

Slot#	Description	Location code	PHB	Max card size
Slot 1	PCIe 8x GX Slot 1	P1-C1 P1-C8	PCIe PHB0	Short
Slot 2	PCIe 8x	P1-C2	PCIe PHB1	Short
Slot 3	PCIe 8x	P1-C3	PCIe PHB3	Long
Slot 4	PCI-X DDR, 64-bit, 266 MHz	P1-C4	PCI-X PHB0	Long
Slot 5	PCI-X DDR, 64-bit, 266 MHz	P1-C5	PCI-X PHB1	Long

2.6.1 System ports

The Service Processor (SP) has two DB9 connectors that are called *system ports*. When the machine is operating, the two system ports become host virtual system ports and are not

general RS232 serial ports, but rather are limited use ports available for specifically supported functions.

The use of the two integrated system ports on the Power 520 is limited to serial connected TTY console functionality and IBM approved call-home modems and other possible IBM service functions. These system ports do not support other general serial connection uses, such as UPS, IBM PowerHA heartbeat, printers, mice, track balls, and so on. If you need serial port function, optional PCI adapters are available; they are described in 2.8.8, "Asynchronous adapters" on page 82.

The integrated console/modem port usage described here is for a Power 520 configured as a single, system-wide partition. When the Power 520 is configured with multiple partitions, the system ports are disabled.

If an HMC is connected, a *virtual serial console* is provided by the HMC (logical device vsa0 under AIX). The system ports are not usable in this case. Either the HMC ports or the system ports can be used, but not both.

Configuration of the two system ports, including basic port settings (such as baud rate), modem selection, and call-home, can be accomplished with the Advanced Systems Management Interface (ASMI).

2.7 Integrated Virtual Ethernet adapter

The POWER6 processor-based servers extend the virtualization technologies introduced in POWER5 by offering the Integrated Virtual Ethernet adapter (IVE). IVE, also named Host Ethernet Adapter (HEA) in other documentation, enables an easy way to manage the sharing of the integrated high-speed Ethernet adapter ports. It has a standard set of features that are part of every POWER6 processor-based server.

The Integrated Virtual Ethernet adapter is a Power 520 standard feature, but you can select from three different combinations of port speed and number of ports. The IVE is provided based upon marketing requirements for improved performance and virtualization for Ethernet. It offers:

► A single controller offering either 2-or 4-port configurations of one of the following:
  – Two 10 Gbps Ethernet ports
  – Four 1 Gbps ports
  – Two 1 Gbps integrated ports
► A low cost Ethernet solution for low-end and mid-range customers or distributed configurations
► Virtual Ethernet resources without the Virtual I/O Server
► Designed to operate at media speeds

The IVE is a physical Ethernet adapter that is connected directly to the GX+ bus instead of being connected to a PCIe or PCI-X bus, either as an optional or integrated PCI adapter. This provides IVE high throughput and low latency. IVE also includes special hardware features to provide logical Ethernet adapters that can communicate to logical partitions (LPAR), reducing the use of POWER Hypervisor (PHYP).

IVE design provides a direct connection for multiple LPARs to share its resources. This allows LPARs to access external networks through the IVE without having to go through an Ethernet bridge on another logical partition, such as a Virtual I/O Server. Therefore, this eliminates the

need to move packets (using virtual Ethernet) between partitions and then through a Shared Ethernet Adapter (SEA) to an Ethernet port. LPARs can share IVE ports with improved performance.

Figure 2-8 provides a general comparison between VIOS SEA and IVE's virtual Ethernet structures. Note that VIOS runs as a partition and IVE runs as part of the system's virtual hardware.
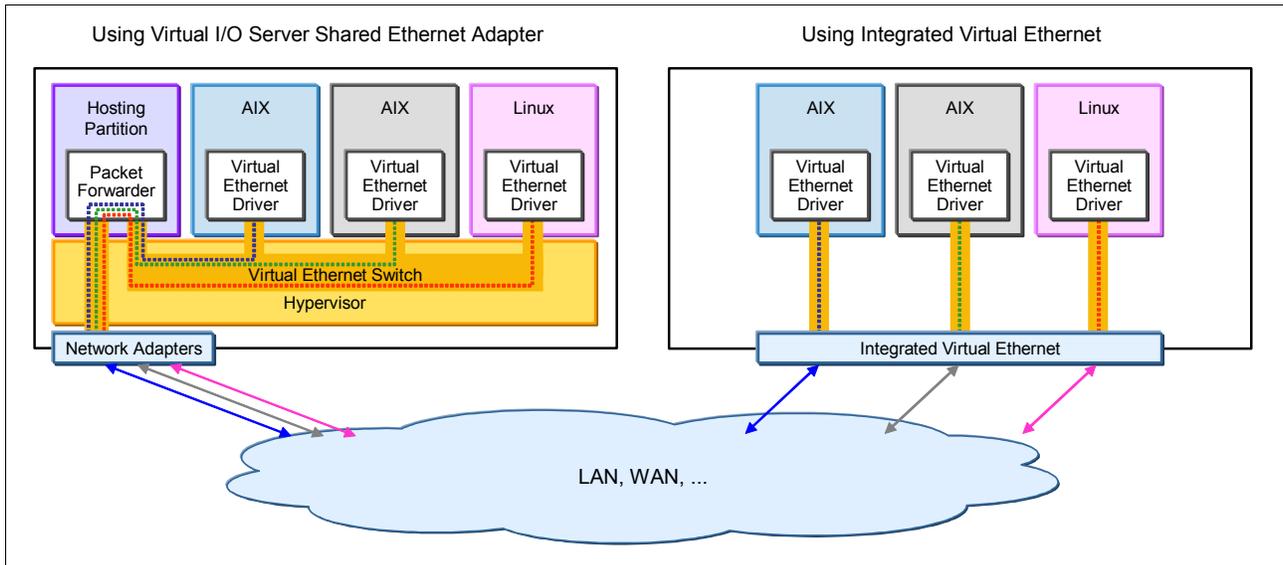
Figure 2-8   Integrated Virtual Ethernet compared to Virtual I/O Server Shared Ethernet Adapter

Depending on the IVE feature ordered, IVE supports 2 or 4 Ethernet ports running at 1 Gbps or 2 ports running at 10 Gbps.

For more information on IVE features you can read *Integrated Virtual Ethernet Adapter Technical Overview and Introduction*, REDP-4340 or review IVE (HEA) information in the IBM Systems Hardware Information Center at:

> http://publib.boulder.ibm.com/infocenter/systems/scope/hw/index.jsp

When an IVE port is expected to have a significant amount of data traffic you should review *System i and System p: Advanced POWER Virtualization Operations Guide*, SA76-0100-03. or later. See the URLs listed in "Other publications" on page 197.

2.7.1 IVE physical ports and system integration

The following sections discuss the physical ports and the features.

Each Power 520 must have one Integrated Virtual Ethernet adapter. The following feature codes are available for a Power 520 server:

► #5623 (standard), Integrated 2-ports 1 Gbps (single controller, 10/100/1000 twisted pair)

– 16 MAC addresses, one port group

► #5613 (optional), Integrated 2-ports 10 Gbps SR[1] (single controller, optical)

– 32 MAC addresses, two port groups

---

[1] 10 Gbps SR (short range) is designed to support short distances over deployed multi-mode fiber cabling; it has a range of between 26 m and 82 m depending on cable type. It also supports 300 m operation over new, 50 µm 2000 MHz·km multi-mode fiber (using 850 nm).

► #5624 (optional), Integrated 4-ports 1 Gbps (single controller, 10/100/1000 twisted pair)

    – 32 MAC addresses, two port groups

The IVE feature code is installed by manufacturing. Similar to other integrated ports, the feature does not have hot-swappable or hot-pluggable capability and must be serviced by a trained IBM System Service Representative.

2.7.2 Feature code port and cable support

All the IVE feature codes have different connectivity options and different cable support.

#5623 and #5624 support:

► 1 Gbps connectivity

► 10 Mbps and 100 Mbps connectivity

► RJ-45 connector

Use the Ethernet cables that meet Cat 5e[2] cabling standards, or higher, for best performance.

#5613 supports:

► Only 10 Gbps SR connectivity

► 62.5 micron multi-mode fiber cable type

    – LC physical connector type

    – 33 meters maximum range

2.7.3 IVE subsystem

Figure 2-9 shows a high level logical diagram of the IVE.
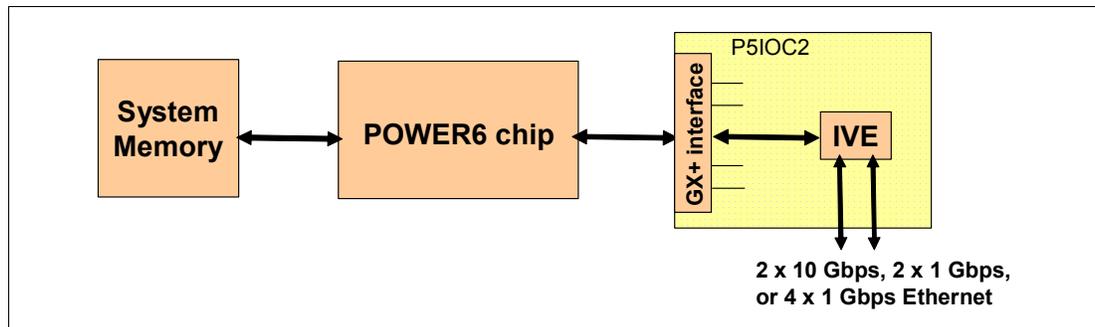
Figure 2-9   IVE system placement

Every POWER6 processor-based server I/O subsystem contains the P5IOC2 chip. It is a dedicated controller that acts as the primary bridge for all PCI buses and all internal I/O devices. IVE major hardware components reside inside the P5IOC2 chip.

The IVE design provides a great improvement in minimizing latency for short packets. Messaging applications such as distributed databases require low latency communication for

---

[2] Category 5 cable, commonly known as Cat 5, is a twisted pair cable type designed for high signal integrity. Category 5 has been superseded by the Category 5e specification.

synchronization and short transactions. The following are IVE characteristics that greatly reduce latency:

► GX+ bus attachment.

► Immediate data in descriptors (reduce memory access).

► Direct user space per connection queuing (OS bypass).

► Designed for up to 3 times throughput improvement over current 10 Gbps solutions.

► Provides additional acceleration functions to reduce host code path length. These include header/data split to help with zero-copy stacks.

► Provides I/O virtualization support so that all partitions of the system can natively take advantage of the previously listed features.

► Allows one 10 Gbps port to replace up to 10 dedicated PCI 1 Gbps adapters in a partitioned system.

One of the key design objectives of the IVE is the integration of up to two 10 Gbps Ethernet ports or four 1 Gbps Ethernet ports into the P5IOC2 chip, with the effect of a low cost Ethernet solution for low-end and mid-range server platforms. Any 10 Gbps, 1 Gbps, 100 Mbps, and 10 Mbps speeds share the same I/O pins and do not require additional hardware or features on top of the IVE card assembly itself. Another key goal is the support of all the state-of-art NIC functionality provided by leading Ethernet NIC vendors.

IVE offers the following functions with respect to virtualization:

► Up to 32 logical ports identified by MAC address

► Sixteen MAC addresses assigned to each IVE port group

► Each logical port can be owned by a separate partition (LPAR)

► Direct data path to LPAR

► Function enablement per LPAR

► Default send and receive queues per LPAR

► Ethernet MIB and RMON counters per LPAR

► VLAN filtering per logical port (4096 VLANs * 32 Logical Ports)

► Internal layer 2 switch for LPAR-to-LPAR data traffic

► Multicast/Broadcast redirection to Multicast/Broadcast manager

IVE relies exclusively on the system memory and system unit processing cores to implement acceleration features. There is not a requirement for dedicated memory, thus reducing the cost of this solution and providing maximum flexibility. IVE Ethernet MACs and acceleration features consume less than 8 mm^2 of logic in CMOS 9SF technology.

IVE does not have flash memory for its firmware but it is stored in the Service Processor flash and then passed to POWER Hypervisor (PHYP) control. Therefore, flash code update is done by PHYP.

2.8 PCI adapters supported in the system unit

POWER6 systems support PCI, PCI-X, PCI-X DDR (Double Data Rate) and Peripheral Component Interconnect Express (PCIe) adapters.

The following subsections discuss the supported adapters under various classifications (LAN, WAN, SCSI, SAS, Fibre Channel (FC), and so forth), and provide tables of orderable feature numbers. The tables indicate whether the feature is supported only on upgrades (cannot be ordered new), the operating systems that support the adapter feature, and whether the adapters can be placed into the PCI slots within the system unit (processor enclosure). A zero (0) value indicates the adapter card is not supported in the system unit.

In most cases a value of zero in the system unit column also implies a supporting IOP card is required for that adapter (IOA) card. Some adapters can run with or without a supporting IOP.

Most of these adapters can also be packaged within a supporting I/O enclosure. However, specific I/O enclosures are not included in these tables. To determine whether an adapter is supported in an I/O enclosure or can be used with a supporting IOP card, review the feature descriptions in any of the following publications:

► *IBM Power 520 and Power 550 (POWER6) System Builder*, REDP-4412.

► Two publications found at the IBM Systems Hardware Information Center Web site:

– Power Systems: PCI adapter placement for machine type 94*nn*

– Power Systems: PCI adapter placement for machine types 82*nn* and 91*nn*

These can be found at:

http://publib.boulder.ibm.com/infocenter/systems/scope/hw/index.jsp

Search for "PC adapter placement."

Note that the 1-core 8203-E4A does not support a GX adapter card for attaching I/O enclosures via either a RIO-2 or 12x loop.

PCIe uses a serial interface and allows for point-to-point interconnections between devices using directly wired interface between these connection points. A single PCIe serial link is a dual-simplex connection using two pairs of wires, one pair for transmit and one pair for receive, and can only transmit one bit per cycle. It can transmit at the extremely high speed of 2.5 GBps, which equates to a burst mode of 320 MBps on a single connection. These two pairs of wires are called a lane. A PCIe link may be comprised of multiple lanes. In such configurations, the connection is labeled as x1, x2, x8, x12, x16, or x32, where the number is effectively the number of lanes.

IBM offers PCIe adapter options for the 8203-E4A, as well as PCI and PCI-extended (PCI-X) adapters. All adapters support Extended Error Handling (EEH). PCIe adapters use a different type of slot than PCI and PCI-X adapters. If you attempt to force an adapter into the wrong type of slot, you can damage the adapter or the slot. A PCI adapter can be installed in a PCI-X slot, and a PCI-X adapter can be installed in a PCI adapter slot. A PCIe adapter cannot be installed in a PCI or PCI-X adapter slot, and a PCI or PCI-X adapter cannot be installed in a PCIe slot.

PCI adapters are also supported in optional I/O enclosures that may be connected to the system via a GX+ adapter loop. Through 2008, PCIe adapters are not supported in available I/O enclosures. They are supported in the Power 520, Power 550, and Power 570 system enclosures (processor enclosure).

For a full list of the adapters that are supported on the systems and for important information regarding adapter placement, see the following PCI adapter placement documents:

► *Power Systems: PCI adapter placement for machine type 94xx*

► *Power Systems: PCI adapter placement for machine types 82xx and 91xx*

These can be found at:

http://publib.boulder.ibm.com/infocenter/systems/scope/hw/index.jsp

Search for "PC adapter placement."

Before adding or rearranging adapters, use the System Planning Tool to validate the new adapter configuration. See the System Planning Tool Web site at:

http://www-03.ibm.com/servers/eserver/support/tools/systemplanningtool/

If you are installing a new feature, ensure that you have the software required to support the new feature and determine whether there are any existing PTF prerequisites to install. To do this, use the IBM Prerequisite Web site at:

http://www-912.ibm.com/e_dir/eServerPrereq.nsf

In the following feature summary tables, note the following:

► Unless otherwise noted, the feature is supported by the AIX, IBM i, and Linux operating systems and release levels listed in 1.16, "Operating system support" on page 51, and on all Power 520 MTMs.

► The focus is on adapters supported in the system unit. Many of these adapters are also supported in external I/O enclosures (attached to the system using a GX+ loop adapter.

Some features supported on a 9408-M25 indicate an IOP is required to support the adapter (IOA). Because the Power 520 system unit does not support an IOP, a RIO-2 I/O enclosure is required. In most cases an equivalent function IOP-less adapter exists and should be used wherever possible.

► Some features are supported for migration (upgrade from a non-POWER6 configuration) purposes and cannot be ordered new. This is noted within the table notes.

► Over time, additional features might be withdrawn from marketing. This will typically be included in announcement letters. To ensure you have the latest information on ordering a feature go to the IBM support Web site. Select the technology (Power in our case) and machine type and then a Planning tab. For example:

http://www.ibm.com/systems/support

– Select **Power**. Select Product machine type 8203-E4A.
– Click **Go**.
– Select the Plan/upgrade tab.

Alternatively, for example for System i models, use this more direct Web address:

http://www.ibm.com/systems/support/i/planning/upgrade/index.html

2.8.1 LAN adapters

To connect a Power 520 to a local area network (LAN), you can use 10/100/1000 Gbps dual-port Integrated Virtual Ethernet with optional 10/100/1000 Mbps quad-port or dual-port 10 Gbps Integrated Virtual Ethernet, one of which must be ordered with the system unit.

Other LAN adapters are supported in the system unit's PCI slots or in I/O enclosures attached to the system using either a RIO-2 technology loop or a 12X technology loop. Table 2-7 on page 78 lists the supported LAN adapters. In the table, A stands for AIX, L for Linux, and i for IBM i. The System Unit Max column refers to the number of adapters that can be supported in the available PCI slots. A value of zero (0) means that adapter card cannot be plugged into one of the available PCI card slots.

Table 2-7 lists the additional LAN adapters that are available in the system unit.

Table 2-7 Available LAN adapters

Feature code	Adapter description	Slot type	Size	System Unit Max	OS support
1954[1]	4-port 10/100/1000 Base-TX	PCI-X	Short	2	A,L
1981[1]	10 Gigabit Ethernet-SR	PCI-X	Short	2	A,L
1982[1]	IBM 10 Gigabit Ethernet-LR	PCI-X	Short	2	A,L
1983[1]	2-port 10/100/1000 Ethernet	PCI-X	Short	2	A,L
2849[2]	10/100 Mbps Ethernet Adapter (Direct SNA support with IOP)	PCI-X	Short	0	i
5706	2-port 10/100/1000 Base-TX	PCI-X	Short	2	A,L,i
5717	4-port 1 Gb Ethernet PCIe 4x	PCIe	Short	3	A,L
5718[3]	10 Gigabit Ethernet -SR	PIC-X	Short	2	A,L
5719[3]	IBM 10 Gigabit Ethernet -LR	PCI-X	Short	2	A,L
5721	10 Gb Ethernet - Short Reach	PCI-X	Short	2	A,L,i
5732	110 GB Ethernet-CX4 PCI Express Adapter	PCIe	Short	2	A,L
5740	4-port 10/100/1000 Ethernet	PCI-X	Short	2	A,L
5767	2-port 1 Gb Ethernet (UTP)	PCIe	Short	3	A,L,i
5768	2-port 1 Gb Ethernet (Fiber)	PCIe	Short	3	A,L,i
5769	10 Gigabit Ethernet-SR PCI Express Adapter	PCIe	Short	3	A, L
5772	10 Gigabit Ethernet-LR PCI Express Adapter	PCIe	Short	3	A,L,i

1. Supported on 8203-E4A, 8204-E8A for migration only. Cannot be ordered new.
2. Note that this code is shared with the GXT135P on AIX and Linux systems.
3. Supported on 8203-E4A, 8204-E8A. Cannot be ordered new.

2.8.2 WAN adapters

Table 2-8 lists the supported WAN adapters.

Table 2-8 Available Wide Area Network (WAN) adapters

Feature code	Adapter description	Slot type	Size	System Unit Max	OS support
2742	PCI Two-Line WAN IOA with IOP	PCI	Short	2	i
2793	PCI 2-Line WAN w/Modem with IOP (non-CIM)	PCI	Short	2	i
2805	4-Line WAN w/Modem with IOP (non-CIM)	PCI	Long	0	i
2806	4-Line WAN w/Modem with IOP (Complex Impedance Matching (CIM))	PCI	Long	0	i
2893	2-Line WAN w/Modem with IOP (non-CIM)	PCIe	Short	0	L,i
2894	2-Line WAN with Modem with IOP (CIM) (IBM i, Linux)	PCIe	Short	0	L,i

Feature code	Adapter description	Slot type	Size	System Unit Max	OS support
5723	(#5723) 2-Port Asynchronous EIA-232 PCI Adapter	PCI	Short	2	A,L
6805	2-Line WAN IOA No IOP	PCI	Short	2	i
6808	4-Line Modem WAN IOA No IOP	PCI	Short	2	i
6834	2-Line WAN w/Modem No IOP (CIM)	PCI	Short	2	i
2943	8-Port Asynchronous Adapter EIA-232/RS-422, PCI bus	PCI	Short	2	A,L
2947	4-port ARTIC960Hx	PCI	Short	2	A
2962	2-Port Multiprotocol PCI Adapter. Includes X.25 packet switched network support. Cannot be ordered new.	PCI	Long	2	A

2.8.3 SCSI and SAS adapters

These adapters enable connection to external SCSI or SAS devices. The adapters available are listed in Table 2-9.

Table 2-9   Available SCSI and SAS adapters

Feature code	Description	Slot	Size	System Unit Max	OS support
1912[1]	DDR Dual Channel Ultra320 SCSI	PCI-X	Short	2	A, L
2749[1]	PCI Ultra™ Mag Media Controller	PCI	Short	0	i
2757[1] (5581, 5591)	Ultra 320 SCSI controller with IOP (up to 757 MB write cache and 1 GB read cache). 5581, 5591 are aux. write cache cards	PCI-X	Short	0	i
2780[1] (5580, 5581, 5583, 5590, 5591)	Ultra 320 SCSI controller with IOP (up to 757 MB write cache/ 1 GB read cache). 5581, 5591 are aux. write cache cards	PCI-X	Short	0	i
5702[1]	PCI-X Ultra Tape Controller	PCI-X	Short	0	i
5712[1]	PCI-X Dual Channel Ultra320 SCSI adapter	PCI-X	Short	0	i
5736	DDR Dual Channel Ultra320 SCSI with IOP	PCI-X	Short	2	A, L, i
5739[1]	PCI-X EXP24 Controller - 1.5 GB Write/1.6 GB Read caches with IOP	PCI-X	Long	0	i
5775	PCI-X Disk/Tape Controller without IOP	PCI-X	Short	2	i
5776	PCI-X DASD Controller-90 MB No IOP	PCI-X	Long	2	i
5777	PCI-X Disk Controller-1.5 GB No IOP	PCI-X	Long	2	i
5778	PCI-X EXP24 Controller - 1.5 GB Write/1.6 GB Read caches without IOP	PCI-X	2 adjct long slots	1 (takes 2 card slots)	i
5782	PCI-X EXP24 Controller - 1.5 GB Write/1.6 GB Read caches without IOP	PCI-X	2 adjct long slots	1 (takes 2 card slots)	i

5806	PCI-X DDR Dual Channel Ultra320 SCSI Adapter - for specific IBM i tape device support (IOP-required)	PCI-X	Long	0	i
5901	PCIe Dual - x4 SAS Adapter	PCIe	Short	3	A, L, i
5902	PCI-X DDR Dual - x4 3 Gb SAS RAID Adapter. Required in pairs.	PCI-X	Short	2	A, L
5903	PCIe Dual - x4 3 Gb SAS RAID Adapter	PCIe	Short	3	A, L
5904	PCI-X DDR 1.5 GB cache SAS RAID Adapter	PCI-X	Long	0	A, L, i
5908	PCI-X DDR 1.5 GB cache SAS RAID Adapter (BSC)	PCI-X	Long	0	A, L, i
5912	PCI-X DDR Dual - x4 SAS Adapter	PCI-X	Short	2	A, L, i
1. Supported on 8203-E4A for migration only. Cannot be ordered new.					

2.8.4 iSCSI adapters

iSCSI adapters in IBM Power 520 (MTM 8203-E4A) provide the advantage of increased bandwidth through the hardware support of the iSCSI protocol. The 1 Gb iSCSI TOE (TCP/IP Offload Engine) PCI-X adapters support hardware encapsulation of SCSI commands and data into TCP, and they transport these commands and data over the Ethernet using IP packets. The adapter operates as an iSCSI TOE. This offload function eliminates host protocol processing and reduces CPU interrupts. The adapter uses a Small form factor LC-type fiber optic connector or a copper RJ45 connector.

Table 2-10 identifies the available iSCSI adapters.

Table 2-10   Available iSCSI adapters

Feature code	Description	Slot	Size	System Unit Max	OS support
1986[1]	1 Gbps SCSI TOE PCI-X on copper media adapter	PCI-X	Short	2	A, L
1987[1]	1 Gbps iSCSI TOE PCI-X on optical media adapter	PCI-X	Short	2	A, L
5713	1 Gbps SCSI TOE PCI-X on copper media adapter	PCI-X	Short	2	A, L, i
5783	1 Gbps Copper iSCSI target HBA	PCI-X	Short	2	i
5784	1 Gbps Optical iSCSI target HBA	PCI-X	Short	2	i
1. Supported on 8203-E4A for migration only. Cannot be ordered new.					

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 and Linux acts as a software iSCSI initiator and allows you 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.

2.8.5 Fibre Channel adapters

The IBM Power 520 (MTM 8203-E4A) supports direct or SAN connection to devices using Fibre Channel adapters. Table 2-11 provides a summary of the available Fibre Channel adapters. The System Unit Max column refers to the number of adapters that can be supported in the available PCI slots. A value of zero (0) means that an adapter card cannot be plugged into one of the available PCI card slots.

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

Supported data rates between the server and the attached device or switch are as follows:

► Distances of up to 500 m running at 1 Gbps
► Distances up to 300 m running at 2 Gbps data rate
► Distances up to 150 m running at 4 Gbps

When you use these adapters with IBM supported Fibre Channel storage switches that support long-wave optics, distances of up to 10 km can run at 1 Gbps, 2 Gbps, and 4 Gbps data rates.

The available adapters are listed in Table 2-11.

Table 2-11   Available Fibre Channel adapters

Feature code	Description	Slot	Size	System Unit Max	OS support
1910[1]	4 Gbps 2-port Fibre Channel	PCI-X	Short	2	A, L
1977[1]	2 Gbps 1-port Fibre Channel Adapter	PCI-X	Short	2	A. L
2787[1]	PCI-X Fibre Channel Disk Controller	PCI-X	Short	0	i
5704[1]	PCI-X Fibre Channel Tape Controller	PCI-X	Short	0	i
5716[1]	2 Gbps 1-port Fibre Channel Adapter	PCI-X	Short	2	A, L
5735	8 Gbps PCIe Dual Port Fibre Channel Adapter	PCIe	Short	3	A, L. i
5749	4 Gbps Dual-Port Fibre Channel PCI-X 2.0 DDR Adapter	PCI-X	Short	2	i
5759	DDR 4 Gbps dual port Fibre Channel	PCI-X	Short	2	A, L
5760	PCI-X Fibre Channel Disk Controller	PCI-X	Short	2	i
5761	PCI-X Fibre Channel Tape Ctlr	PCI-X	Short	2	i
5773	1-port 4 Gbps Fibre Channel	PCIe	Short	3	A, L
5774	4 Gbps PCIe Dual Port Fibre Channel Adapter	PCIe	Short	3	A, L, i
1. Supported on 8203-E4A for migration only. Cannot be ordered new.					

2.8.6 Graphic accelerators

The Power 520 supports up to four graphics adapters. Table 2-12 identifies the available graphic accelerators. They can be configured to operate in either 8-bit or 24-bit color modes. These adapters support both analog and digital monitors.

Table 2-12   Available graphic accelerators

Feature Code	Description	Slot	Size	System Unit Max	OS Support
1980[1]	GXT135P Graphics Accelerator	PCI-X	Short	2	A, L
2849[1]	GXT135P Graphics Accelerator	PCI-X	Short	2	A, L
5748	GXT145 Graphics Accelerator	PCIe	Short	3	A, L
1. Supported on 8203-E4A for migration only. Cannot be ordered new.					

Note: These adapters are not hot-pluggable.

2.8.7 Cryptographic adapters

IBM offers POWER6 cryptographic coprocessor and secure-key cryptographic accelerator cards in several features. The coprocessor functions are targeted to banking and finance applications. Financial PIN processing and Europay, MasterCard, and Visa (EMV) credit card functions are provided. EMV is a standard for integrated-chip based credit cards. The secure-key accelerator functions are targeted to improving the performance of Secure Sockets Layer (SSL) transactions.

Table 2-13 lists the available cryptographic adapters. The System Unit Max column refers to the number of adapters that can be supported in the available PCI slots. A value of zero (0) means that an adapter card cannot be plugged into one of the available PCI card slots.

Table 2-13   Available cryptographic adapters

Feature code	Description	Slot	Size	System Unit Max	OS support
4764	Cryptographic Coprocessor (FIPS 4) and secure-key cryptographic accelerator functions in a single PCI-X card.	PCI-X	Short	2	A, L, i
4806[1]	PCI-X Cryptographic Accelerator	PCI-X	Short	0	i
1. Supported on Power 520 for migration only. Cannot be ordered new.					

2.8.8 Asynchronous adapters

Asynchronous PCI-X adapters provide connection of asynchronous EIA-232 or RS-422 devices. If you have a cluster configuration or high-availability configuration and plan to connect the IBM System p servers using a serial connection, the use of the two integrated system ports is not supported. Use one of the features listed in Table 2-14 on page 83. In the table, A stands for AIX and L stands for Linux. The System Unit Max column refers to the number of adapters that can be supported in the available PCI slots.

Table 2-14   Asynchronous PCI-X adapters

Feature Code	Description	Slot	Size	System Unit Max	OS Support
2943[1]	8-Port Asynchronous Adapter EIA-232/RS-422	PCI-X	Short	2	A, L
5723	2-Port Asynchronous IEA-232 PCI Adapter	PCI-X	Short	2	A, L
5785	4-Port Asynchronous IEIA-232 PCIe Adapter	PCIe	Short	--	A, L
1. Supported on 8203-E4A for migration only. Cannot be ordered new					

2.8.9 Special purpose adapters/coprocessors, IOPs, and associated features

The major PCI-based adapters that you can configure within a POWER6 system unit or attached I/O enclosure when you build a supported configuration are described in the previous sections.

These include most, but not all supported adapters. We cannot list all supported adapters, cabling features, and devices attached to the adapters described in this publication.

There remain some other hardware components that may be supported uniquely by AIX, Linux for Power, and IBM i.

Examples of adapters or devices we do not include in this publication are:

► Cryptographic coprocessors and cryptographic accelerators
► Tape devices and the IBM System Storage Tape Libraries
► IBM i's twinaxial workstation controller support for local 5250 workstation support

To get the complete list of adapters and devices supported by each MTM and operating systems, and more, contact your IBM marketing or service representative. Consider also reviewing the following IBM Web sites and documents:

► IBM Announcement Letters and Sales Manual pages for the specific system and MTM

► The IBM prerequisites Web site (system or MTM, feature number, and supporting operations system release levels):

   http://www-912.ibm.com/e_dir/eServerPrereq.nsf/

► IBM System Storage Interoperation Center (SSIC) Web site:

   http://www-03.ibm.com/systems/support/storage/config/ssic/displayesssearchwithoutjs.wss?start_over=yes

► For System i-based attachments and adapters, *IBM System i Overview: Models 515, 525, 550, 570, 595, and More,* REDP-5052, found at:

   http://www.ibm.com/redbooks

2.9 System unit storage

This section expands on certain capabilities and features unique to the IBM Power 520 (MTM 8203-E4A) system unit itself. For example, each system unit features one SAS DASD controller with up to six hot-swappable 3.5 inch SAS disk bays or eight hot-swappable 2.5 inch (SFF) SAS disk bays and one hot-pluggable, slim-line media bay per enclosure. Only the new SAS DASD hard disk drives or solid state drives are supported within the system unit.

IBM i requires its disks to be protected. Therefore, if IBM i is to use up to eight disks in the system unit enclosure, at least two disks are required (IBM i mirroring support) with a minimum of three disks if RAID 5 is used.

There is no requirement that the system unit have any disks installed.

Table 2-15 lists the DASD backplane options.

Table 2-15   SAS DASD backplane options

Feature code	Description	Supported OS
8308	DASD/Media Backplane for 3.5 DASD/SATA DVD/Tape (without External SAS Port)	AIX / Linux
8310	DASD/Media Backplane for 3.5 DASD/SATA DVD/Tape (with External SAS Port)	AIX / Linux / IBM i
8346[1]	DASD/Media Backplane for 2.5 DASD/SATA DVD/Tape (with External SAS Port)	AIX / Linux / IBM i
1. For IBM i, IBM i 6.1 is required.		

If feature code 3586, 3587, 3647, 3648, 3649, 3658, 3677, or 3678 (3.5 inch SAS disks) is selected, then the following rules apply:

► Either feature code 8308 or 8310 must be selected.

► Feature code 1843 (Op Panel Cable for Deskside System with 3.5 inch DASD) or 1877 (Op Panel Cable for Rack Mount Drawer with 3.5 inch DASD) must be selected:

– Feature code 1843 must be selected with deskside cover set 7112 or 7113.
– Feature code 1877 must be selected with rack mount bezel 7200 or 7201.

If feature code 1882, 1883, 1884, 1890, 1909 (2.5 inch SAS disks) is selected, then the following rules apply:

► Feature code 8346 must be selected.

► Feature code 1843 (Op Panel Cable for Deskside System with 3.5 inch DASD) or 1877 (Op Panel Cable for Rack Mount Drawer with 3.5 inch DASD) must be selected:

– Feature code 1843 must be selected with deskside cover set feature code 7112 or 7113.
– Feature code 1877 must be selected with rack mount bezel feature code 7200 or 7201.

The 8310 and 8346 offer split backplane support. Split backplane support is available to AIX and Linux but is not supported by IBM i. This feature provides the ability to split the backplane and allocate a SAS controller for each half of the internal disks. Split DASD support hardware requirements are as follows:

► High function DASD backplane (8310 or follow-on)
► SAS Cable, DASD Backplane (Split) to Rear Bulkhead (3670)
► PCI-X Dual Channel Ultra320 SCSI Adapter (5912 or follow-on)
► External SAS cable

Note that 8310 is the default, base function backplane with external SAS port. The 8346 is the optional, higher capacity DASD backplane with an external SAS port and 2.5 inch SFF SAS Disk connectivity.

There are currently two supported drawers:

► The 5886 EXP 12S SAS Disk Drawer, which is an expansion drawer with 12 SAS storage slots. The 5886 supports up to 12 hot-swap SAS disk drives in mega-pack carriers. See 2.12, "5886 EXP 12S SAS DASD Expansion Drawer" on page 97 for more details.

► The 5802 EXP PCIe 12X I/O Drawer, which is an expansion drawer with 18 SAS SFF storage slots. The 5802 supports up to 18 hot-swap SFF SAS disk drives in 9 pack carriers.

The SAS RAID Enablement feature (5679) provides the internal SAS enclosure with RAID levels 0, 5, 6, and 10 function and Auxiliary cache using backplane pluggable daughter cards. The daughter cards contain the RAID function and 175 MB of Write Cache. The Auxiliary daughter card plugs into a special slot on the planar and provides a battery power pack and redundant 175 MB write cache memory for the daughter card.

> **Note:** To take advantage of advanced features, whether for functions with current or future availability, select 8310 and 8346 over the 8308 to avoid the need to replace the backplane at a later time.
>
> When configuring the 8310 or 8346, you can connect only one IBM Storage EXP 12S expansion drawer. For more disk capacity, order the 5904 or 5908 instead.

Refer to 2.12, "5886 EXP 12S SAS DASD Expansion Drawer" on page 97 for information about 8203-E4A connection to the EXP 12S SAS Disk Drawer attached to the system unit through 5679 and the 8310 or 8346 backplane.

Currently there are several physical capacity SAS disk drives that can be placed within the system unit, the EXP12S, or the EXP PCIe I/O Drawer. Their capacities are dependent upon the operating-system-unique formatting required on the disk. Table 2-16 lists the 8203-E4A and disk drive feature numbers that each bay can contain.

Table 2-16   SAS Disk drive feature code description: 8203-E4A and AIX, IBM i disks

Feature code	Description	OS support	Supported in		
			CEC	**5802**	**5886**
1882	SAS HDD SFF 146.8 GB 10,000 rpm	A, L	Yes	Yes	-
1883	SAS HDD SFF 73.4 GB 15,000 rpm	A, L	Yes	Yes	-
1884[1]	SAS HDD SFF 69.7 GB 15,000 rpm	i	Yes	Yes	-
1890	SAS SSD SFF 69.7 GB	A, L	Yes	No	-
1909[1]	SAS SSD SFF 69.7 GB	i	Yes	No	-
3586	SAS SSD 3.5" 69.7 GB	A, L	No	-	Max 8
3587	SAS SSD 3.5" 69.7 GB	i	No	-	Max 8
3647	SAS HDD 3.5" 146.0 GB 15,000 rpm	A, L	Yes	-	Yes
3648	SAS HDD 3.5" 300.0 GB 15,000 rpm	A, L	Yes	-	Yes
3649	SAS HDD 3.5" 450.0 GB 15,000 rpm	A, L	Yes	-	Yes
3658[2]	SAS HDD 3.5" 428.0 GB 15,000 rpm	i	Yes	-	Yes
3677	SAS HDD 3.5" 139.5 GB 15,000 rpm	i	Yes	-	Yes
3678[2]	SAS HDD 3.5" 283.7 GB 15,000 rpm	i	Yes	-	Yes
1. Requires IBM i 6.1 2. If used as IBM i load source device, you must use IBM i 6.1. - indicates not applicable.					

2.9.1 Embedded SAS disk controller

The six or eight disk bays are run by the imbedded disk controller. Because the controller is "owned" by one partition, the owning partition needs to virtualize storage controlled by the imbedded disk controller to any secondary partition that needs disk storage from these disks. Many client environments can improve disk performance of the six drives by using the optional 5679 175 MB write cache.

The 5679 does not consume a PCI slot. It is two cards. The first card provides RAID support and a 175 MB write cache. The second card is the auxiliary 175 MB write cache and battery to help avoid single points of failure that might cause extended outages. With the 5679 175 MB write cache feature, RAID 5 or RAID 6 disk protection can be implemented for the six or eight drive slots in the system unit. Mirroring protection is available through the operating system. RAID 5 requires a minimum of three drives, and RAID 6 requires four. The same imbedded controller that runs the disk drives also runs the SAS tape slot and the slimline DVD drive in the system unit. The 5679 write cache is not used in tape or DVD operations.

Figure 2-10 illustrates the 5679 card with battery.
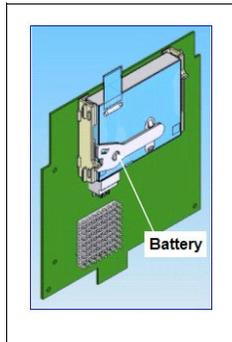
Figure 2-10   5679 RAID auxiliary cache card, with battery

2.9.2 Disk and disk controller protection rules

The AIX, Linux, and IBM i operating systems provide disk data protection under mirroring and various levels of RAID protection. Some protection is performed by the operating system and some by the disk adapter (controller) supporting the disks. Each operating system provides some protection methods uniquely and some identically.

2.9.3 Internal optical devices

This section describes the optical devices that are supported by the media bay within the system unit enclosure. A slimline optical device is required. The system has a slim media bay that can contain an optional DVD-ROM or an optional DVD-RAM. Table 2-17 lists the media feature codes available.

Table 2-17   Media bay features

Feature Code	Description	OS Support
5743	SATA Slimline DVD-ROM Drive	A, L, i
5762	SATA Slimline DVD-RAM Drive	A, L, i
Feature not orderable, planned for end of support November 2009.		

2.9.4 Internal tape devices

This section describes the tape devices that are supported by the media bay within the system unit enclosure.

A half-high tape device is not required. The system has a half-high bay that can contain a tape drive. Table 2-18 lists the media feature codes and the cartridge feature codes available. The System Unit Max column refers to the number of adapters that can be supported in the available PCI slots. If a tape device is installed in the half-high media bay, 3655 SAS HH Cable must be selected.

Table 2-18   Media bay and cartridge features

Feature code	Description	System Unit Max	OS support
5619	80/160 GB DAT 160 SAS Tape Drive	1	A, L, i
5746	Half High 800 GB / 1.6 TB LTO4 SAS Tape Drive; Write/Read Ultrium4 (LTO-4), Write/Read Ultrium3 (LTO-3), Read Ultrium2 (LT02) formats.	1	A, L, i
5907	Half High 36/72 GB 4 mm DAT72 SAS Tape Drive	1	A, L, i
5689	DAT 160 Tape Cartridge (5X), 80/160 GB	1	A, L, i
5747	Ultrium4 - 800 GB LTO-4 Tape Cartridge (5X)	1	A, L, i

2.9.5 Split backplane support

IBM Power Systems 520(8203-E4A) and 550(8204-E8A) offer split backplane support. Split backplane support is available to AIX and Linux but is not supported by IBM i. This feature provides the ability to split the backplane and allocate a SAS controller for each half of the internal disks. With the announcement of 2.5" SFF support on the Power Systems the following two configurations are possible:

► With the 3.5 inch disks you can allow three of the six SAS disk bays to be accessed via an external SAS port on the back of the p6 520 or p6 550. When this occurs, then three disk bays (accessed via the PCI-X SAS controller) can be assigned to one LPAR and three disk bays (accessed via the integrated SAS controller) can be assigned to another LPAR.

► With the 2.5 inch SFF disks you can allow four of the eight SAS disk bays to be accessed via an external SAS port on the back of the p6 520 or p6 550. When this occurs, then four disk bays (accessed via the PCI-X SAS controller) can be assigned to one LPAR and four disk bays (accessed via the integrated SAS controller) can be assigned to another LPAR.

Split DASD support requirements are as follows:

► First half of drive bays always run by the embedded controller and optionally augmented by write cache and RAID enabler #5679.

► Second half of drive bays when split run by the following SAS adapter (PCI-X or PCIe):
   – Zero write cache PCI-X #5900/5912 or PCIe-#5901
   – 1.5 GB write cache PCI-X #5904 (This option makes the most sense only for high performance SSD given the price of this adapter.)
   – #5902/#5903 medium cache adapter option not announced for 520/550 split backplane

► High function DASD backplane (#8310/#8346).

► SAS Cable, DASD Backplane (Split) to Rear Bulkhead #3670 for 520 and #3669 for 550 (different cables are required because the 550 is deeper than the 520 and needs longer

cable). This SAS cable connects the DASD Backplane with External SAS Port (#8310/#8346) to the rear bulkhead of the system.

► External SAS cable(#3679). This cable provides connection to a slot in the SAS controller.

> **Note:** Split backplane is incompatible when attaching a #5886 EXP12S Disk Drawer to a CEC SAS port.

Figure 2-11 shows the schematic of the split backplane connectivity.
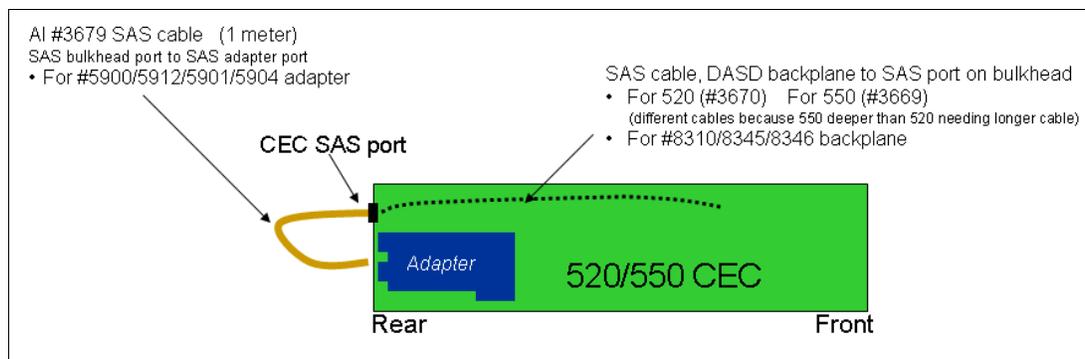
Figure 2-11   Split backplane support for 520 and 550

2.10 Drawer and tower attachment

As discussed in 2.5, "I/O capabilities" on page 70, the system unit has five PCI I/O expansion slots and up to two GX+ I/O loop adapter slots (0 on 1-core, 1 on 2-core, and 2 on 4-core).

There are two PCIe 8X short-length slots, one PCIe 8X full-length slot, and two PCI-X DDR long slots. The PCIe short slot 1 is shared with the first GX+ slot. A second GX+ slot is available on a 4-core configuration and does not share a PCI slot.

In 2.8, "PCI adapters supported in the system unit" on page 75, we discuss the hardware features that can be plugged into the PCIe and PCI-X DDR slots, as well as features that can be plugged into I/O enclosures that are supported on either a RIO-2 or 12X I/O loop using loop cabling connecting the I/O enclosure's GX+ adapter with the system unit's GX+ adapter.

If an adapter requiring a supporting IOP is needed or more PCI slots are needed than can be contained by the system unit, the optional GX adapters (not supported on 1-core configurations) can be installed, allowing up to two I/O loops with up to four (12X loops) or six (RIO-2 loops) I/O enclosures per loop to be ordered and configured. By the term *I/O enclosures* we include the following "container" terms that are associated with a specific orderable feature.

► I/O drawers
► I/O expansion units
► I/O towers

Existing model configurations that can be upgraded to POWER6 configurations have a set of I/O enclosures that have been supported on RIO-2 (HSL-2) loops for a number of years. Most continue to be supported on POWER6 and POWER6+ models.

In 1.3.5, "I/O enclosures attachment using 12x or RIO-2 I/O loop adapters" on page 17, we provide tables and additional descriptions of the supported I/O enclosures and whether they connect using RIO-2 or 12X. Each loop can have only RIO-2 or 12X I/O enclosures.

Table 2-19 lists the available system unit GX+ adapter types and their feature numbers. The System Unit Max column refers to the number of adapters that can be supported in the available PCI slots. A value of zero (0) means that an adapter card cannot be plugged into one of the available PCI card slots. On configurations supporting up to two GX+ adapter loops, you can have all loops 12X or RIO-2, or one of each technology.

Table 2-19   GX adapters

Feature code	Description	System Unit Max	OS support
5609	GX Dual-port 12X (DDR) channel attach adapter. The 5609 adapter provides up to twice the data rate capability as the 5616 12X adapter.	0, 1 (min 4-core required)	A, L, i
5614	Dual port RIO-2 I/O hub	0, 1, 2	A, L, i
5616	GX dual-port 12X (SDR) channel attach adapter	0, 1, 2	A, L, i

The AIX, LInux, and IBM i operating systems support a defined set of I/O enclosures and I/O adapters on the loop that is connected to the GX adapter. The enclosures supported are unique based upon RIO-2 or 12X loop technology and the operating system.

2.11 IBM TotalStorage Expansion 24 Disk Drawer overview

The IBM TotalStorage EXP24 Expandable Storage disk enclosure was known by different names and by MTM or feature numbers to users experienced with System p and System i, as follows:

► System p: *7031 Model D24 (I/O Drawer)* or *T24 (I/O Tower)* (that is, the 7031-D24 or 7031-T24)

► System i: *5786 TotalStorage EXP24 Disk Drawer* or *5787 TotalStorage EXP24 Disk Tower*

Table 2-20   Availability and withdrawal summary

GEO	Type Model	Announced	Available	Marketing withdrawn	Service discontinued
WW	7031-D24	2005/10/04	2005/10/14	2009/02/27	-
WW	7031-T24	2005/10/04	2005/10/14	2009/05/29	-
WW	5787	2007/02/16	2007/02/16	2008/04/08	-
WW	5786	2007/02/16	2007/02/16	-	-

Note: The unified models of POWER Systems should now use #5786 as the feature code to order Exp 24 Disk drawer for IBM i only. AIX/Linux SCSI disks are no longer orderable on EXP24.

These I/O enclosures were identical regardless of feature code differences. The enclosed Ultra320 (LVD) SCSI disk drives are physically the same, but have different feature numbers

because of different physical capacities based upon the formatting required by the using operating systems. This disk storage enclosure device provides more than 3 TB of disk storage in a 4U rack-mount unit. Whether as 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 accommodated in any of the four six pack disk drive enclosures are 141.12 GB 15K rpm drives. Each of the four six pack disk drive enclosures can be attached independently to an Ultra320 SCSI or Ultra320 SCSI RAID adapter.

For highly available configurations, a dual bus repeater card (#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 can be connected to form a single Ultra320 SCSI bus of 12 drives.

These disk units can be packaged in up to four independent groups of six disk units. Each group of six disk units is referred to as a *six pack*.

Each six pack with one or more disks is enabled by either a #5741 Expansion 24 6 Disk Slot Enabler or a #5742 Expansion 24 6/12 Disk Slot Enabler. Each six pack must be driven by a port on a supporting SCSI disk controller located outside the #5786.

> **Note:** Each *disk slot enabler* is also referred to as a *SCSI repeater* in IBM documentation. User documentation uses the terms *repeater* and *enabler* interchangeably. One repeater is required for each six pack regardless of the type of repeater.

The port on the disk controller is connected to either a #5741 or a #5742 using a SCSI cable. One to four disk slot enablers (repeaters) are required, depending on the number of six packs that are populated with disk units.

The #5741 Expansion 24 6 Disk Slot Enabler is termed a *single SCSI repeater* and the #5742 Expansion 24 6/12 Disk Slot Enabler is termed a *dual SCSI repeater*. The #5742 can support a single six pack or up to two six packs (up to 12 disks), when functioning as a dual repeater.

When functioning as a dual repeater, a #5742 must be connected to the supported disk controller port. It then can be daisy-chain connected, using a SCSI cable, to either another #5741 or #5742 (the second six pack). This second repeater must not be connected to any other repeater or disk controller port.

> **Notes:**
> - The #5786 does not use HSL cables and does not have SPCN connections.
> - When the EXP is connected to a single I/O adapter (IOA) through its enabler, then up to 36 DASD can be attached to each IOA. Minimum drives in a RAID 5 set is four and the maximum is 18. Therefore, a minimum of two RAID sets with 36 drives are attached to the IOA. We recommend that you configure up to 24 disks within a single EXP24 from a best performance viewpoint.

The disk controller features supporting attachment of a #5786 TotalStorage Expansion 24 Disk Drawer include the #5736, #5737, #5739, #5775, #5776, #5778, #5781, and #5782.

Specific disk unit feature numbers are used to identify disk units that will be placed in the #5786. These disk units are physically the same disk units as used in other System i system units and I/O towers or drawers. Using separate feature codes allows IBM configuration tools to better understand their placement. #4328 is a unified feature code number which is

available to order on #5786, and other supported disk units include the #1266, #1267, #1268, #1269, #1292, #1293, #1294, #1295, #1296, #1297, #1298, and #1299.

#5786 EXP24 disk configuration: The #5786 and #5787 EXP24 disk configurations can support a larger number of disk drives with fewer disk controllers (IOAs) and therefore use fewer PCI slots than was previously required with older disk controllers. From an attachment standpoint, a single disk controller (IOA) and a single EXP24 allow up to 24 disks to be physically attached.

When one and a half EXP24 enclosures are attached to a single controller (three SCSI buses), up to 36 disks can be attached. However, for performance reasons, 24 disks per high-speed controller is usually the practical limit and is inferred as the maximum in most EXP24-based documentation. By comparison, the previous maximum per disk controller was 20 disk drives in a mirroring environment (where four SCSI buses could be used), or 15 to 18 disk drives in a RAID 5 environment with auxiliary write cache support.

Because the EXP24 enclosure is organized into four sets of up to six drives, each set of six disks can be attached to the same or different disk controller IOAs. This flexibility can be of significant value when configuring small logical partitions (LPARs).

2.11.1 EXP24 Disk Drawer front and rear views

Figure 2-12 shows the front view of the #5786 TotalStorage Expansion 24 Disk Drawer.
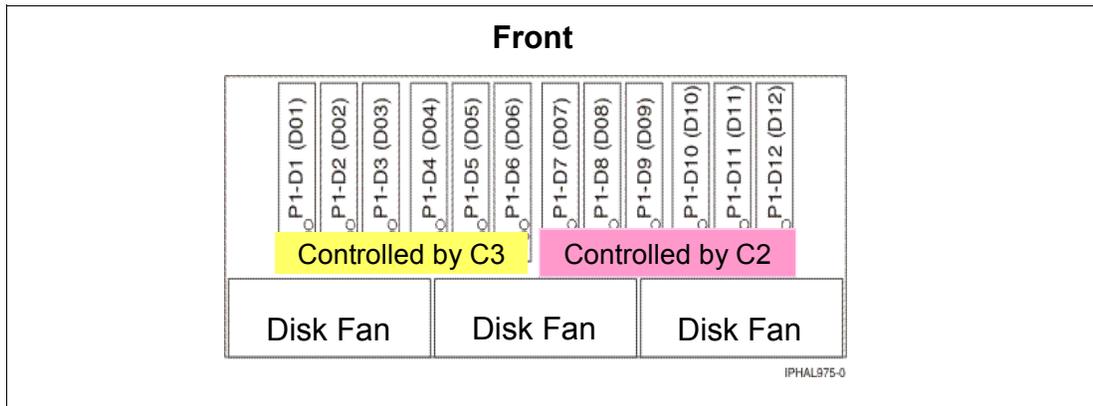

Figure 2-12   Front view of the #5786

Figure 2-13 shows the back view of the #5786 TotalStorage Expansion 24 Disk Drawer. In the back view, C2, C3, C4, and C5 are slots for the #5741 and #5742 SCSI repeaters.
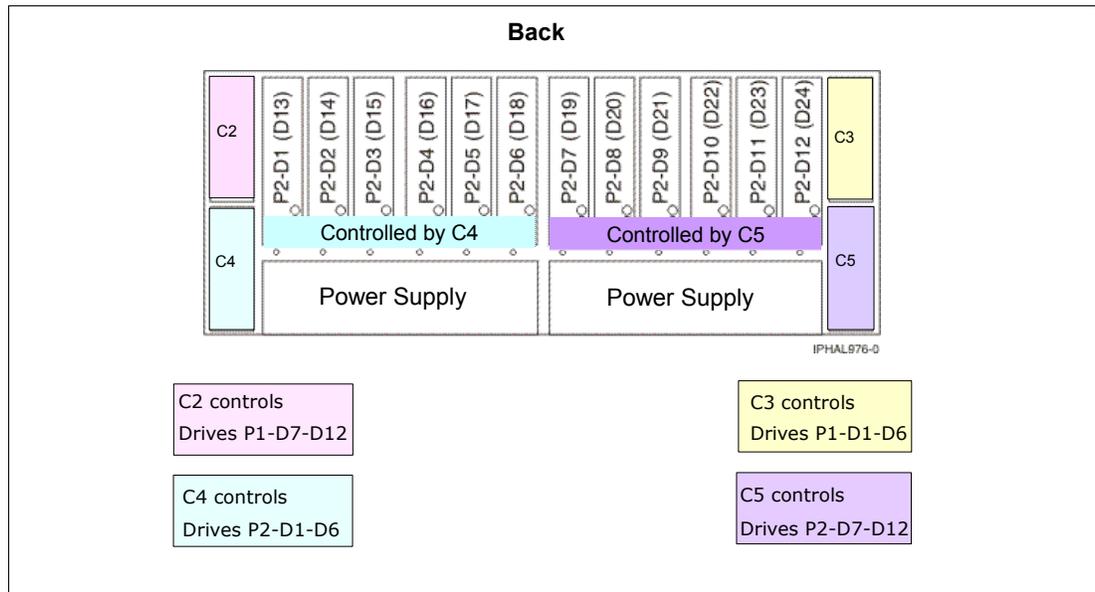

Figure 2-13   Rear view of the #5786

Note: On tower models (#5787), slots C3 and C5 are located on the top and C2 and C4 are located on the bottom. Thus, the C3 is the top left, the C5 is the top right, the C2 is the bottom left, and C4 is the bottom right when viewed from the rear. Follow the location codes when placing repeater cards.

2.11.2 Adapter cards for connecting EXP24 Disk Drawer

Table 2-21 lists the adapter cards to connect the EXP24 Disk Drawer.

Table 2-21   Adapter cards for connecting EXP24 Disk Drawer

	Type	E8A	E4A
#1912 (withdrawn)	IOP-less	◔	◔
#5736	With IOP	◔	◔
#5737	With-IOP		
#5739	With-IOP		
#5775	IOP-less		
#5776	IOP-less		
#5778	IOP-less	◔	◔
#5781	With IOP		
#5782	IOP-less	◔	◔

2.11.3 5741 and 5742 Disk Slot Enabler placement and cabling recommendations

A repeater (also referred to as an *enabler*) is required whenever there are disk drives in the slots that are controlled by that repeater. The repeater slots can be populated in several ways:

► All #5741 single repeaters (allowing up to six disks per disk controller port).

► All #5742 dual repeaters (not recommended because it is more expensive and unnecessary).

► For greatest flexibility, use two #5741 single repeaters and two #5742 dual repeaters.

 The #5741/#5742 placement affects the disk device SCSI addressing and the number of potential LPAR load source disk units.

► A combination of #5741 single and #5742 dual repeaters (allowing up to 12 disks per controller port).

 – When using a combination of #5742 dual and #5741 single repeaters, we recommend that you place the #5742 dual repeaters in C3 and C4. This placement keeps the disk device SCSI addressing consistent without change, but limits the available LPAR load source disk units to two.

 – When connecting (daisy-chaining) repeaters, the SCSI cabling should go from the disk controller port to a #5742 dual repeater and then to a #5741 single repeater.

 – The repeaters should be connected (daisy-chained) so that a pair of repeaters controls the disk slots in either the front or rear half of the #5786/#5787 (C3 to C2 and C4 to C5).

The SCSI ports on the #5741/#5742 repeaters are *not* labelled. Figure 2-14 shows a representation of the repeaters as viewed from the rear. It shows only the cabling between SCSI repeaters. A complete working configuration for controlling up to 12 disks also requires a SCSI cable connection from a SCSI port on a disk controller to a port on the #5742 dual repeater. In this example, that port on the #5742 is labelled A.
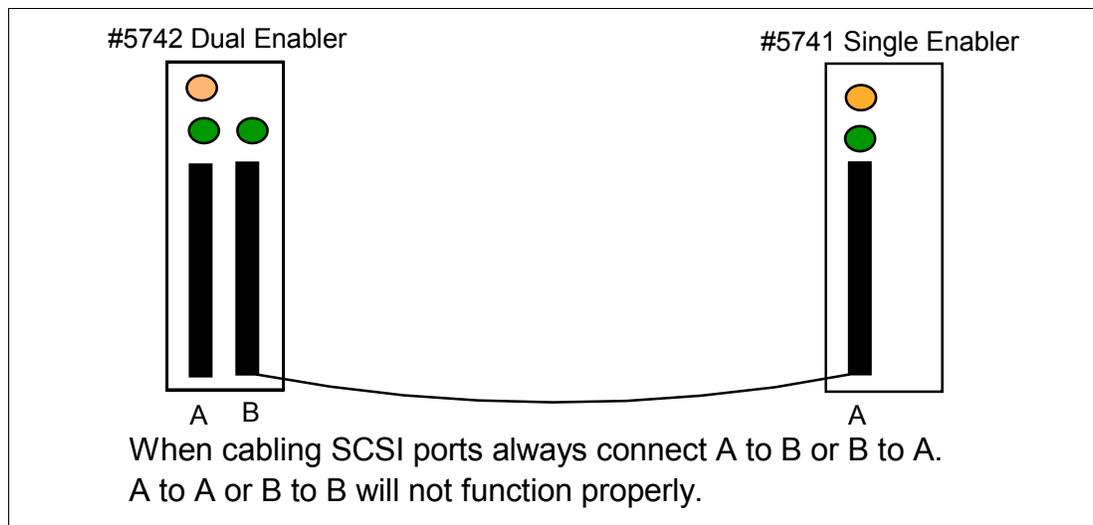

Figure 2-14   Logical cabling view between EXP24 enabler (repeater) adapters

2.11.4 Load source drive considerations in #5786 for logical partitions with 8203-E4A and 8204-E8A

Each IBM i LPAR requires a load source disk unit. The server uses the load source to start the LPAR. Each LPAR has specific supported slot placements for its load source disk depending on the type of system unit or expansion unit where the load source is installed. A specific connection to a disk IOA is then required to control the load source disk unit for each LPAR.

> **System Planning Tool (SPT):** The information provided here does not replace the SPT. Use this information as a resource with the SPT output. Its purpose is to assist you in the load source placement for your IBM i LPARs.

The load source drive requires a SCSI device address of 1, 2, 3, or 4 and must be connected to SCSI port 0 of a disk controller. Under certain circumstances, the SCSI device addresses in the EXP24 are A, B, C, D, E, and F. Therefore, care is needed to ensure addressing is correct when attempting to have more than two load sources contained in a single EXP24.

Figure 2-15 shows the SCSI device addressing and the effect of using a #5742 dual repeater or a #5741 single repeater and the effects of cabling between repeaters.
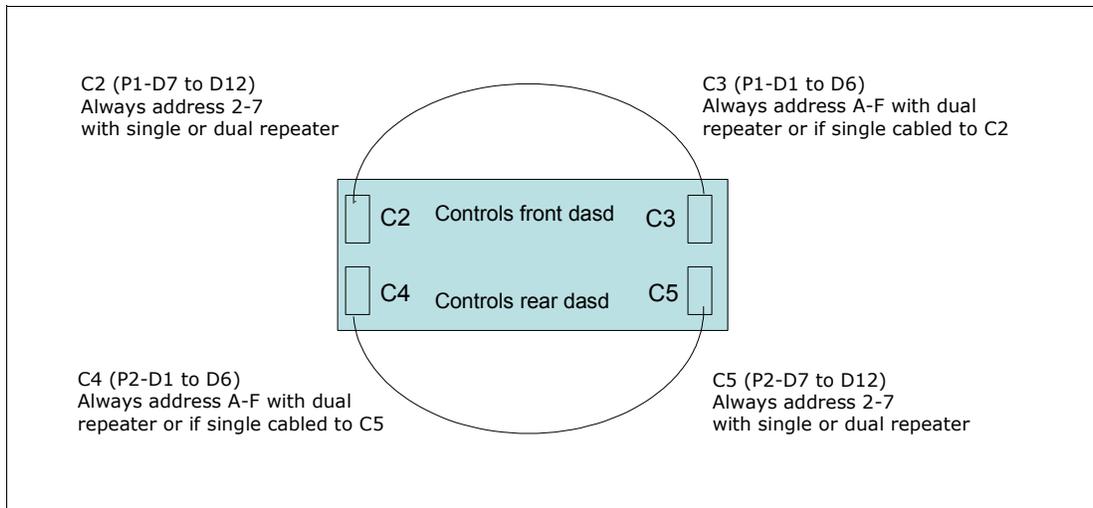

Figure 2-15 Card slot device addressing rules using #5741 single and #5742 dual repeaters

The SCSI addressing for DASD positions P1-D7 to P1-D12 and P2-D7 to P2-D12 does not change. The positions are always SCSI address 2, 3, 4, 5, 6 and 7. This means that the first three positions in the right six pack viewed from the front (P1-D7, D8, D9), and the first three positions in the right six pack viewed from the rear (P2-D7, D8, D9), can always be load source candidates.

P1-D1 to P1-D6 and P2-D1 to P2-D6 device positions change to addresses A, B, C, D, E, and F in two cases. They change whenever a #5742 dual repeater controls them or if a #5741 single repeater controlling them is cabled to a #5742 dual repeater. Addresses A, B, C, D, E, and F are in hexadecimal format. These same addresses can be displayed in decimal format (10, 11, 12, 13, 14, 15) in some IBM i displays such as Hardware Service Manager or error logs.

2.11.5 Load source rules for #5786

Consider the following load source rules for #5786:

► The load source disk must be controlled by the SCSI bus port 0 of the load source disk unit controller.

► P1-D1, P1-D2, or P1-D3 can contain the load source disk only if slot C3 contains a #5741 single repeater card that is connected to SCSI port 0 of the disk controller.

► P1-D7, P1-D8, or P1-D9 can contain the load source if the #5741 in C2 is connected to SCSI port 0 on the disk controller or if the #5741 in C2 is cabled to the #5742 dual repeater in C3 and the #5742 in C3 is connected to SCSI port 0 on the disk controller.

► P2-D1, P2-D2, or P2-D3 can contain the load source disk only if slot C4 contains a #5741 single repeater card that is connected to SCSI port 0 of the disk controller.

► P2-D7, P2-D8, or P2-D9 can contain the load source if the #5741 in C5 is connected to SCSI port 0 on the disk controller or if the #5741 in C5 is cabled to the #5742 dual repeater in C4 and the #5742 in C4 is connected to SCSI port 0 on the disk controller.

2.11.6 #5741 and #5742 SCSI repeater card placement recommendations

Consider the following SCSI repeater card placement recommendations:

► To have four load source disks available in a single #5786/#5787, the #5786/#5787 must contain four #5741 single repeaters or #5742 dual repeaters in C2 and C5 and #5741 single repeaters in C3 and C4.

► To keep SCSI addressing consistent with no changes, but limit the available load source disks to two instead of four, then the #5786/#5787 should have a #5741 single repeater in C2 and C5 and #5742 dual repeaters in C3 and C4.

► When daisy-chaining SCSI repeaters, connect C2 to C3 and connect C5 to C4. This way, the repeaters in C2 and C3 control the disks in the front of the #5786/#5787 and the repeaters in C4 and C5 control the disks in the rear.

2.11.7 IBM i large write/read cache adapters that support the EXP24

IBM i supports several adapters that have a large write cache (390 MB up to 1.5 GB compressed) and read cache (415 MB up to 1.6 GB compressed). These have a second (auxiliary) write cache card which duplicates the primary write cache data for data recovery purposes.

These adapters all report to the system as a specific adapter card CCIN value and an associated CCIN value for the auxiliary write cache card. There are versions of these adapters specific to supporting the EXP24. The EXP24 support CCIN values are 571F (primary adapter and write cache card) and 575B (auxiliary write cache card).

These EXP24 adapter cards are two physical cards that are firmly connected to each other and require two adjacent PCI-X slots. The primary card provides three Ultra320 SCSI ports or buses for the attachment of disk drives located in a #5786 TotalStorage EXP24 Disk Drawer/Tower.

The two cards, identified on the system by a single primary adapter card CCIN (571F), and the auxiliary write cache card CCIN 575B, are ordered as a single feature code out of several orderable feature codes that include these same physical cards. Each of these orderable feature numbers (codes) indicate to the IBM ordering system a unique surrounding hardware configuration.

Orderable features #5739 and #5778 are physically the same adapter cards but have different feature numbers to denote to IBM configuration tools whether an IOP is required. The #5739 indicates that an IOP is used.

Orderable feature numbers #5781 or #5782 and #5799 or #5800 are also the same adapter cards. However, the #5781 (with IOP) or #5782 indicates that the adapter is placed in a double-wide blind swap cassette, and the #5799 (with IOP) or #5800 indicates that the adapter is placed in a system unit and has a light pipe inserted into the feature.

Figure 2-16 depicts the SCSI port addressing on the double-wide (2 physical cards) disk controller (adapter) supporting the EXP24 enclosure.


A- Battery cover

B- SCSI port 2

C- SCSI port 0

D- SCSI port 1
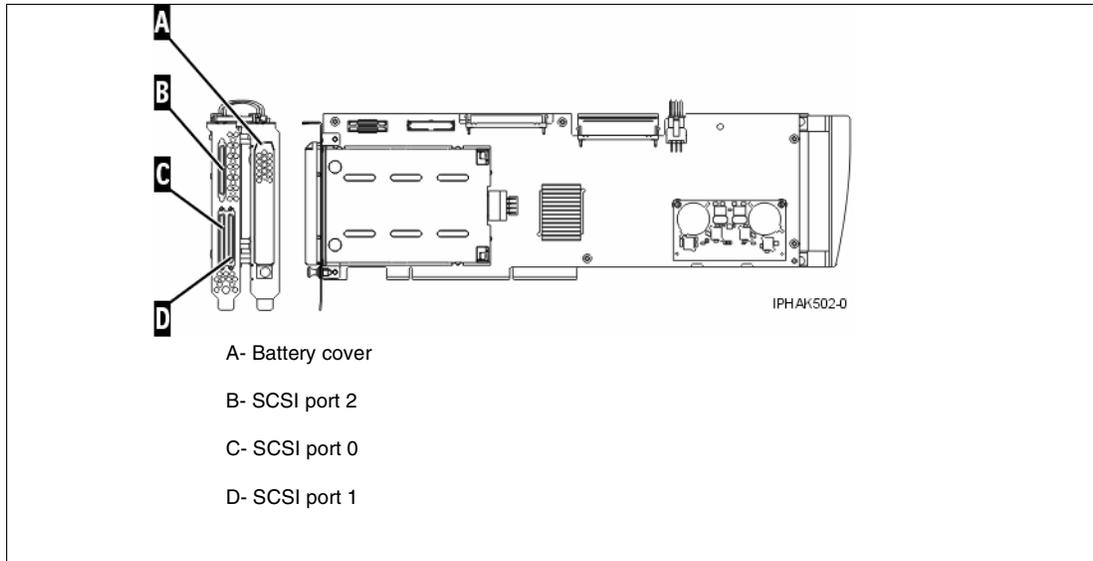
Figure 2-16   EXP 24 ports and batter cover

The IBM Systems Hardware Information Center contains additional information about the EXP24 using both System i (#5786 or #5787) and System p (#7031-D24 or #7031-T24) disk enclosure feature numbers. Refer to the Hardware Information Center on the Web at:

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

Search the Information Center using the feature numbers #5786, #5787, #7031-D24, or #7031-T24. Add the word *repeaters* to the search to find line drawings and instructions in addition to the information that we provide in this paper.

2.11.8 SCSI disk drives on 12X and RIO-2 loops

Figure 2-17 depicts the flexibility of a single EXP24 with disks attached to two different technology loops. Each six pack is shown connected to a supporting adapter within a 12X I/O enclosure and a RIO-2 I/O enclosure.
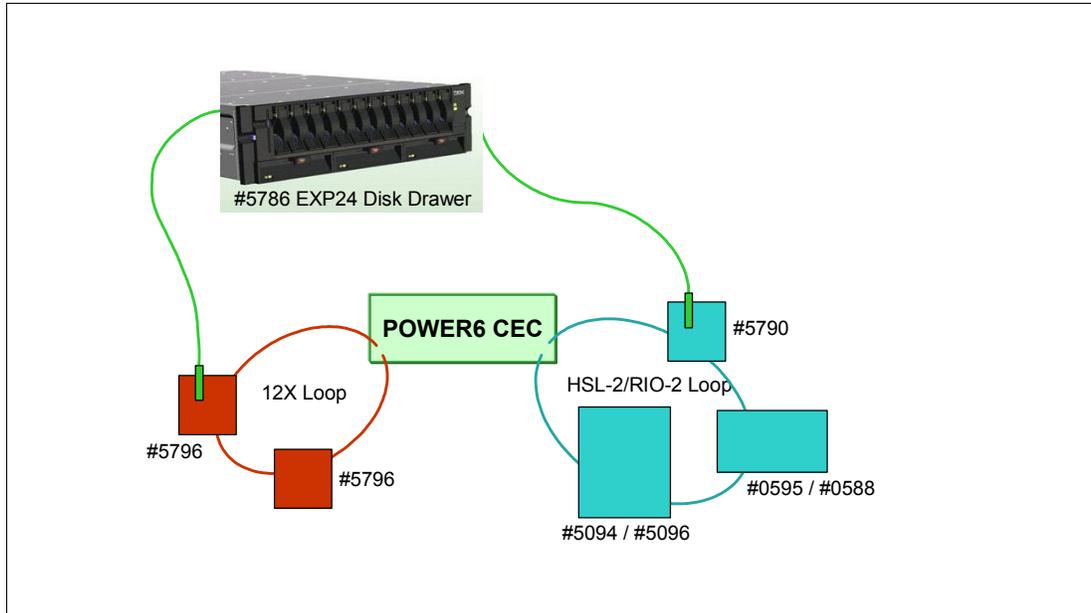

Figure 2-17   EXP24 connection

2.12 5886 EXP 12S SAS DASD Expansion Drawer

The IBM System Storage EXP 12S is an expansion drawer with 12 SAS Storage Slots. The SAS disks are accessible from the front of the 5886. This EXP 12S takes up a 2 EIA units space in a 19 inch rack. The enclosure attaches to a host server using the appropriate external SAS cables connected to supporting controllers (adapters). 5886 EXP 12S SAS DASD Expansion Drawer characteristics include:

► Modular SAS disk expansion drawer
► Up to twelve 3.5 in. SAS disk drives
► Redundant hot-plug power and cooling with dual line cords
► Redundant and hot-swap SAS expanders

The following terms are commonly used when discussing this expansion drawer:

► *SAS Enclosure*

   The 19 inch EIA rack drawer that contains two expanders, two power supplies, a midplane, and up to 12 SAS disk drives.

► *Expander*

   A device that is a part of the SCSI service delivery subsystem and facilitates the communication between SAS devices and the controller. Within the 5886 SAS enclosure, the expander is called the Enclosure Service Manager (ESM)

► *SES Device*

   The SCSI Enclosure Service Device which is contained within the expander.

2.12.1 5886 EXP 12S SAS DASD Expansion Drawer physical description

Table 2-22 lists the 5886 EXP 12S SAS DASD Expansion Drawer attributes.

Table 2-22   5886 attributes

Dimension	One CEC drawer
Height	45 mm (1.7 in.)
Width	440 mm (17.3 in.)
Depth (including front bezel)	735 mm (28.9 in.)
Weight	17.7 kg (120 lb)

2.12.2 5886 EXP 12S SAS DASD Expansion Drawer operating environment

Table 2-23 lists the general system specifications of the system unit.

Table 2-23   5886 specifications

Description	Range (operating)
Operating temperature	10°C to 38°C (50°F to 100.4°F)
Relative humidity	20% to 80% (allowable) 40% to 55% (recommended)
Maximum dew point (operating)	21°C (69.8°F)
Noise level	6.5 bels idle / 6.6 bels operating
Operating voltage	100-240 V ac at 50-60 Hz
Maximum power consumption	700 watts (maximum)
Maximum power source loading	0.740 kVa (maximum)
Maximum thermal output	2,382 BTU[a]/hr (maximum)
Maximum altitude	2134 m (7000 ft)

a. British Thermal Unit (BTU)

2.12.3 5886 front and rear views

Figure 2-18 on page 99 shows the front view of the 5886 EXP 12S SAS DASD Expansion Drawer.

Location Code	Component		Location Code	Component
Un-P1-D1	Disk Drive 1		Un-P1-D7	Disk Drive 7
Un-P1-D2	Disk Drive 2		Un-P1-D8	Disk Drive 8
Un-P1-D3	Disk Drive 3		Un-P1-D9	Disk Drive 9
Un-P1-D4	Disk Drive 4		Un-P1-D10	Disk Drive 10
Un-P1-D5	Disk Drive 5		Un-P1-D11	Disk Drive 11
Un-P1-D6	Disk Drive 6		Un-P1-D12	Disk Drive 12
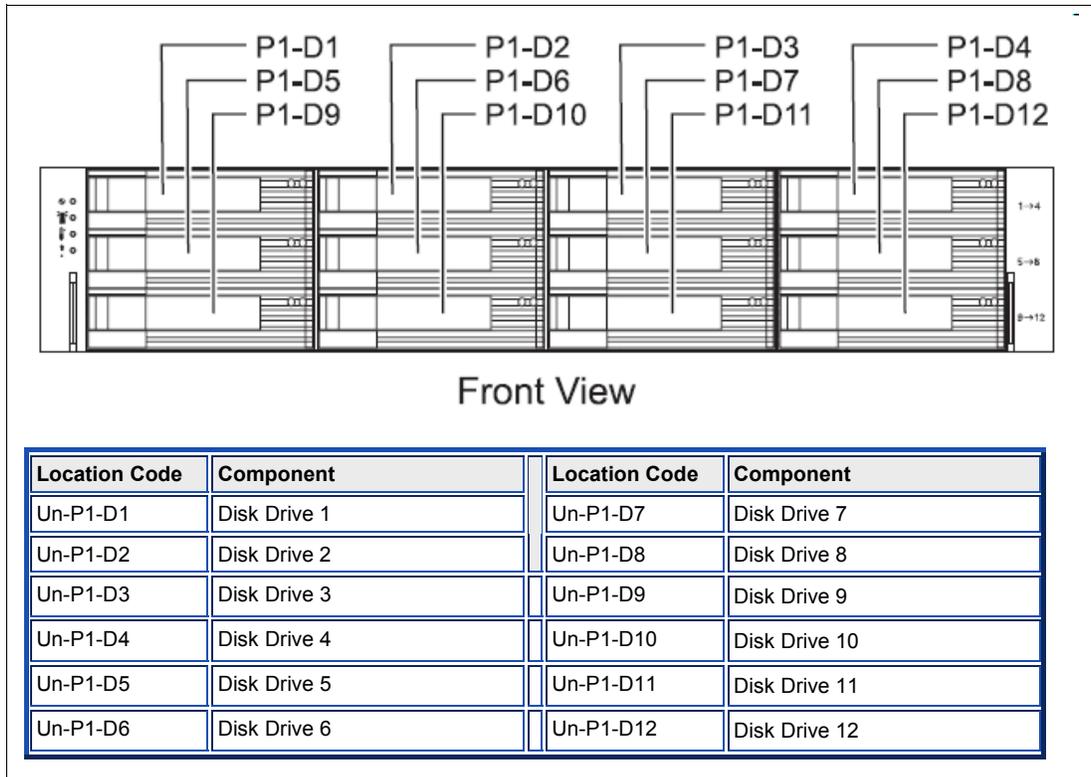
Figure 2-18   Front view 5886

Figure 2-19 shows the rear view of the 5886 EXP 12S SAS DASD Expansion Drawer.

| Location Code	Component		Location Code	Component
U*n*-C1	ESM A		U*n*-E1	Power Supply
U*n* -C2	ESM B		U*n*-E2	Power Supply
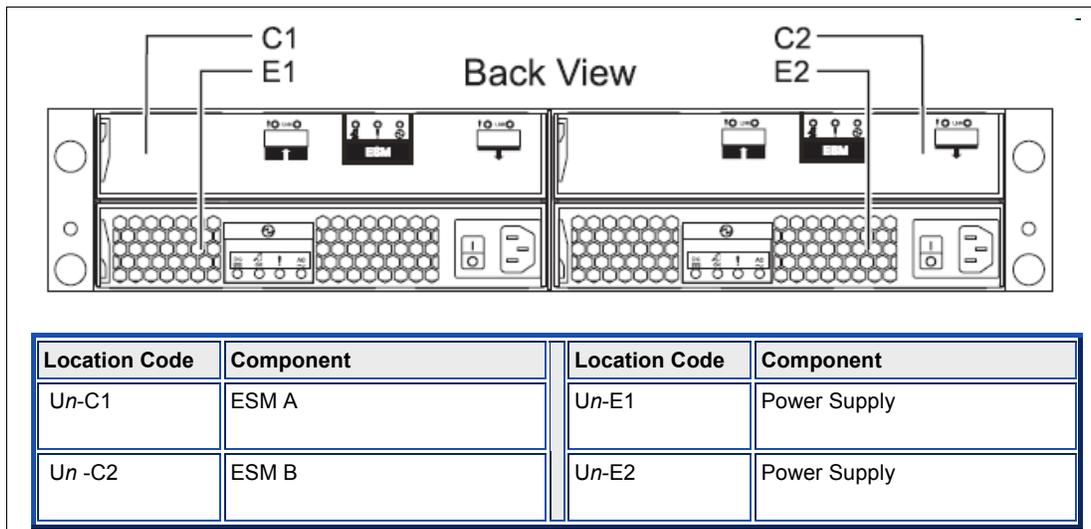
Figure 2-19   Rear view 5886

2.12.4 Host system specific information

The following sections provide more details about the EXP 12 support according to the machine types that are determined by the ordered primary operating system.

Attached to 8203-E4A and 8204-E8A

Note: New disk drives are announced continuously. In addition, older disk drives are removed from marketing. You should periodically review recent announcement letters to determine the currently marketed and supported disk devices, in all supported I/O enclosures.

Table 2-24 lists the drives supported for IBM i, AIX, and Linux.

Table 2-24   SAS disk drive feature code description: 8203-E4A and AIX, IBM i disks

Feature code	Description	OS support
3586	SAS SSD 3.5" 69.7 GB	AIX/Linux
3587	SAS SSD 3.5" 69.7 GB	IBM i
3647	SAS HDD 3.5" 146.0 GB 15,000 rpm	AIX/Linux
3648	SAS HDD 3.5" 300.0 GB 15,000 rpm	AIX/Linux
3649	SAS HDD 3.5" 450.0 GB 15,000 rpm	AIX/Linux
3658[1]	SAS HDD 3.5" 428.0 GB 15,000 rpm	IBM i
3677	SAS HDD 3.5" 139.5 GB 15,000 rpm	IBM i
3678[1]	SAS HDD 3.5" 283.7 GB 15,000 rpm	IBM i
1. If used as IBM i load source device, you must use IBM i 6.1.		

The EXP 12S drawer must be mounted in a 19 inch rack, such as the IBM 7014-T00 or 7014-T42.

To connect an EXP 12S SAS drawer to an 8203-E4A or 8204-E8A system unit, an additional adapter is needed, or the #8346 backplane configured. Table 2-25 provides the current list of available adapters.

Table 2-25   SAS adapters

Feature	Description	OS support
5679	The SAS RAID Enablement feature provides the internal SAS enclosure with 0, 5, 6, and 10 RAID function and 175MB of Write Cache/Auxiliary cache via backplane pluggable daughter cards. Requires Media Backplane #8346	AIX/Linux and IBM i
5901	PCIe Dual-x4 SAS Adapter	AIX/Linux and IBM i 6.1 only
5902	The PCI-X DDR Dual - x4 3 Gb SAS RAID Adapter is an excellent solution for high-performance applications requiring two adapters. PCI-X DDR Dual - x4 3 Gb SAS RAID Adapter supports RAID 0, 5, 6, and 10. Two #5902s provide for mirrored write cache data.	AIX/Linux
5903	(#5903) - PCIe 380MB Cache Dual - x4 3 Gb SAS RAID Adapter used for SAS disks and SAS SSD disks. The #5903 is installed in pairs, allowing for redundancy of the write cache. If the #5903 pairing is broken, then write cache is disabled.	AIX/Linux

Feature	Description	OS support
5904	PCI-X DDR 1.5 GB Cache SAS RAID Adapter. Auxiliary write cache and concurrent battery maintenance are provided. The adapter provides RAID 0, 5, 6, and 10 for AIX and Linux. Under IBM i OS, mirroring and data spreading is provided by the operating system and RAID 5 and 6 is provided by the adapter.	AIX/Linux and IBM i
5908	PCI-X DDR 1.5 GB Cache SAS RAID Adapter blind swap cassette (BSC) #5908 is same card as #5904 but a Gen-3 blind swap cassette used in enclosures such as the 19-inch #5790, 5796, or Power 570 CEC.	AIX/Linux and IBM i
5912	PCI-X DDR Dual - x4 SAS Adapter supports RAID 0 (with mirroring) and 10.	AIX/Linux and IBM i
5922	Feature 5922 must be added for every instance of a non-paired SAS RAID adapter #5902. It identifies a specific high availability configuration supported by AIX or Linux which has one #5902 on one system and the paired #5902 located on a second system.	AIX/Linux
5923	Feature 5923 must be added for every instance of a non-paired SAS RAID adapter #5903. It identifies a specific high availability configuration supported by AIX or Linux which has one #5903 on one system and the paired #5903 located on a second system.	AIX/Linux

Depending on the required configuration a different set of cables is needed to connect the EXP 12S drawer to the system or drawer. These cables are identified in Table 2-26.

Table 2-26   SAS connection cables

Feature	Description
3652	SAS cable (EE) drawer to drawer 1 m
3653	SAS cable (EE) drawer to drawer 3 m
3654	SAS cable (EE) drawer to drawer 3 m
3661	SAS Cable (X) Adapter to SAS Enclosure, Dual Controller/Dual Path 3 m
3662	SAS Cable (X) Adapter to SAS Enclosure, Dual Controller/Dual Path 6 m
3663	SAS Cable (X) Adapter to SAS Enclosure, Dual Controller/Dual Path 15 m
3668	SAS Cable, DASD Backplane to Rear Bulkhead (only for 8204-E8A)
3669	SAS Cable, DASD Backplane (Split) to Rear Bulkhead (only for 8204-E8A)
3670	SAS Cable, DASD Backplane (Split) to Rear Bulkhead (only for 8203-E4A)
3674	SAS Cable, DASD Backplane to Rear Bulkhead (only for 8203-E4A)
3679	SAS Cable (AI)- 1 m
3684	SAS Cable (AE) Adapter to Enclosure, single controller/single path 3 m
3685	SAS Cable (AE) Adapter to Enclosure, single controller/single path 6 m
3686	SAS Cable (YI) System to SAS Enclosure, Single Controller/Dual Path 1.5 m
3687	SAS Cable (YI) System to SAS Enclosure, Single Controller/Dual Path 3 m
3691	SAS cable (YO) adapter to SAS enclosure 1.5 m
3692	SAS cable (YO) adapter to SAS enclosure 3 m

Feature	Description
3693	SAS cable (YO) adapter to SAS enclosure 6 m
3694	SAS cable (YO) adapter to SAS enclosure 15 m

A typical base configuration is using a server machine and a single attached drawer, as shown in Figure 2-20.


YO-Cable

SAS Adapter

IPHAD706-2
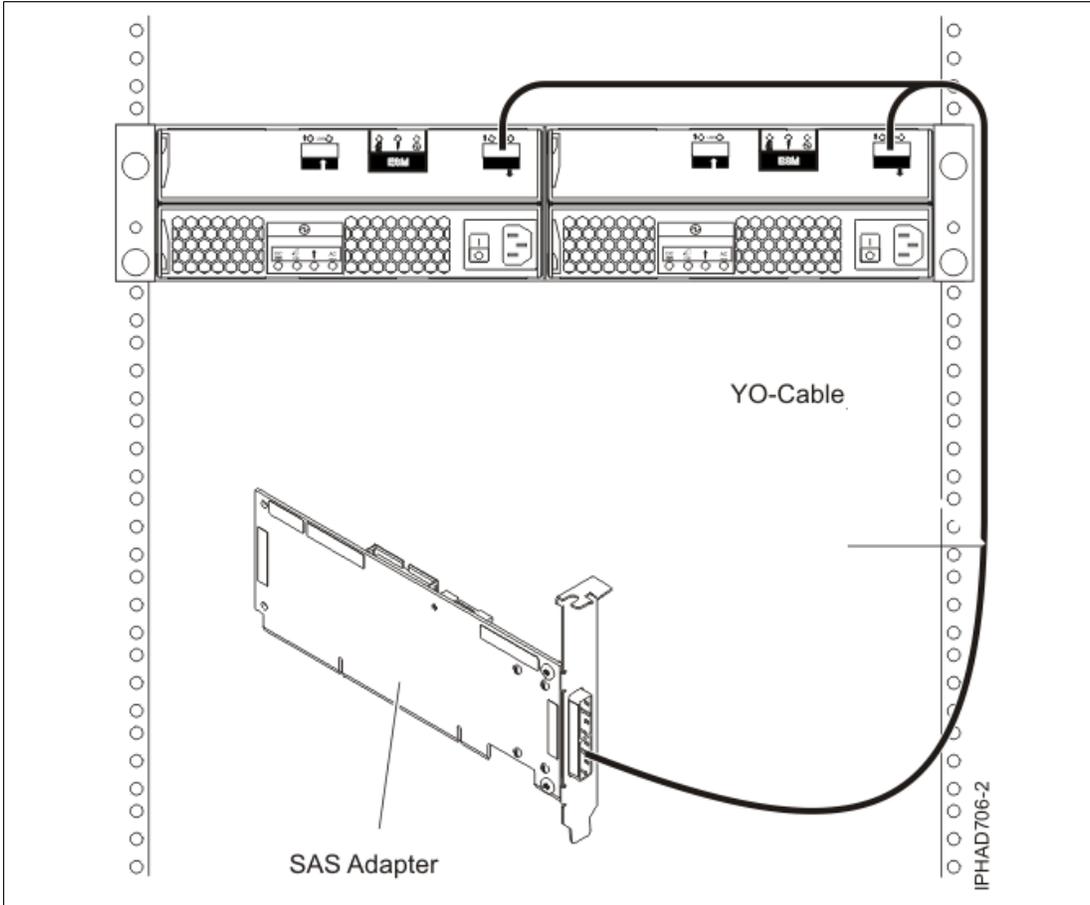
Figure 2-20   Base configuration of one SAS drawer

A maximum configuration using four EXP 12S drawers on one adapter feature is shown in
Figure 2-21.
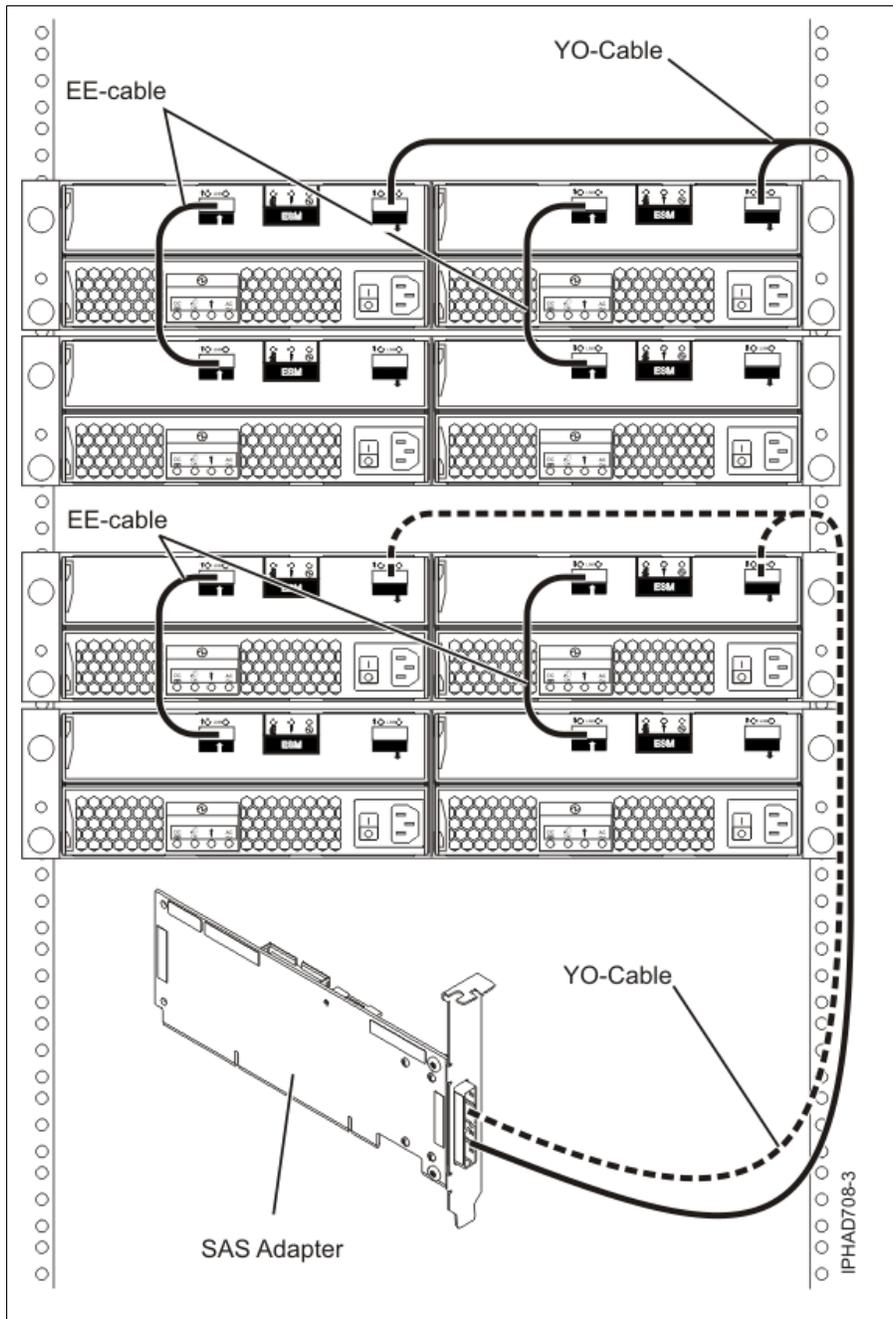
*Figure 2-21   Maximum attachment of EXP 12S on one adapter*
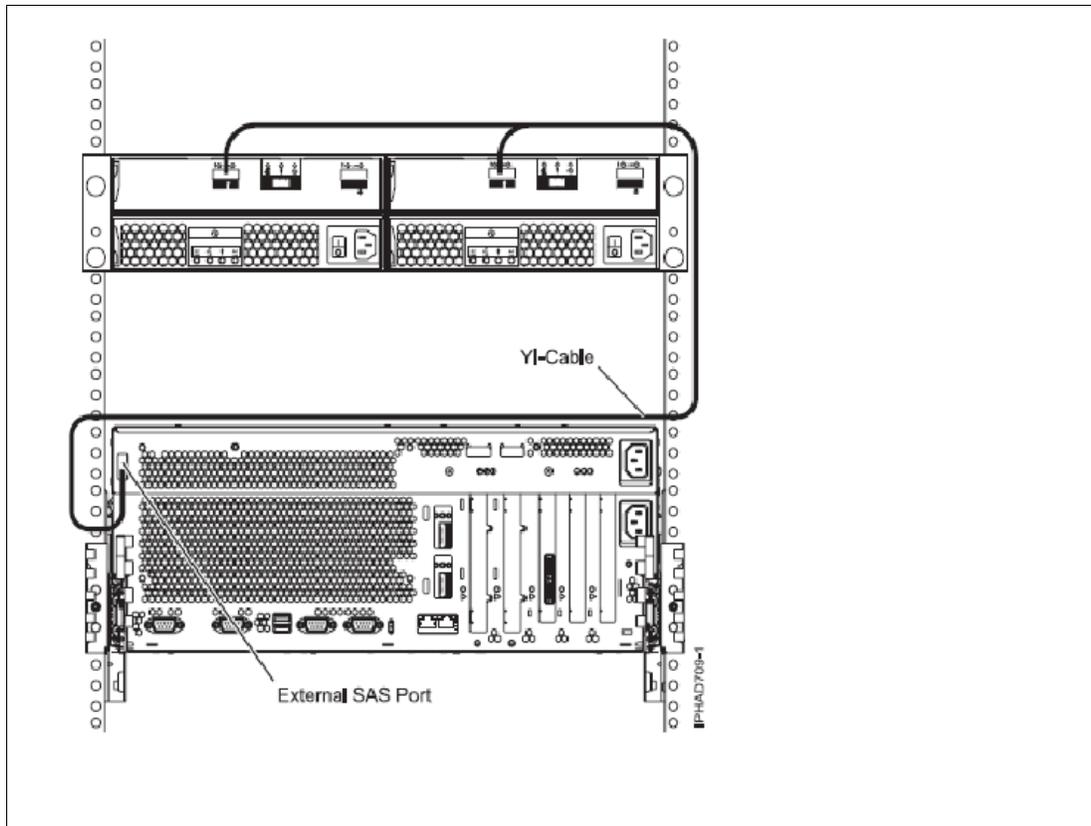
Figure 2-22   Single enclosure attached with YI cable to external SAS port

---

Note:

- ► The one core 8203-E4A (#5633 processor) does not support #5886 EXP 12S SAS Expansion Drawer.
- ► A maximum of 24 drawers per two core 8203-E4A or 8204-E8A is achievable with SAS adapters in remote I/O drawers. Four #5886 can be supported per #5900. Two #5886 per SAS Adapter is recommended for redundant configurations.
- ► The YI cables #3687 (1.5 m) and #3687 (3 m) are used to connect to the port on the #8346 backplane.

Attached to the new SAS Controller #5902 and #5903

The following figure shows the schematic to connect two adapters using the x-cable. Key features of this configuration are:

- ► #5902 or #5903 are always paired to provide redundancy.

- ► The adapters are linked together via SAS X cable and write cache contents are mirrored. If one adapter fails or is disconnected, contents of the write cache are written out and write cache is disabled until pairing is restored. Usually significant performance impact is possible without write cache, especially if running RAID-5 or RAID-6.

- ► Maximum of 48 SAS disk drives per pair of adapters.

- ► Maximum of 24 SAS drives per paired #5902 or per paired #5903 port using two #5886 linked together using EE SAS cables.

- ► There might be performance considerations if you try to attach this many drives.

- Maximum of 8 Solid State Drives per #5903 adapter pair.
- #5903 might throttle SSD performance with this many SSDs.

Figure 2-23 shows splitting the drawer between two adapters.
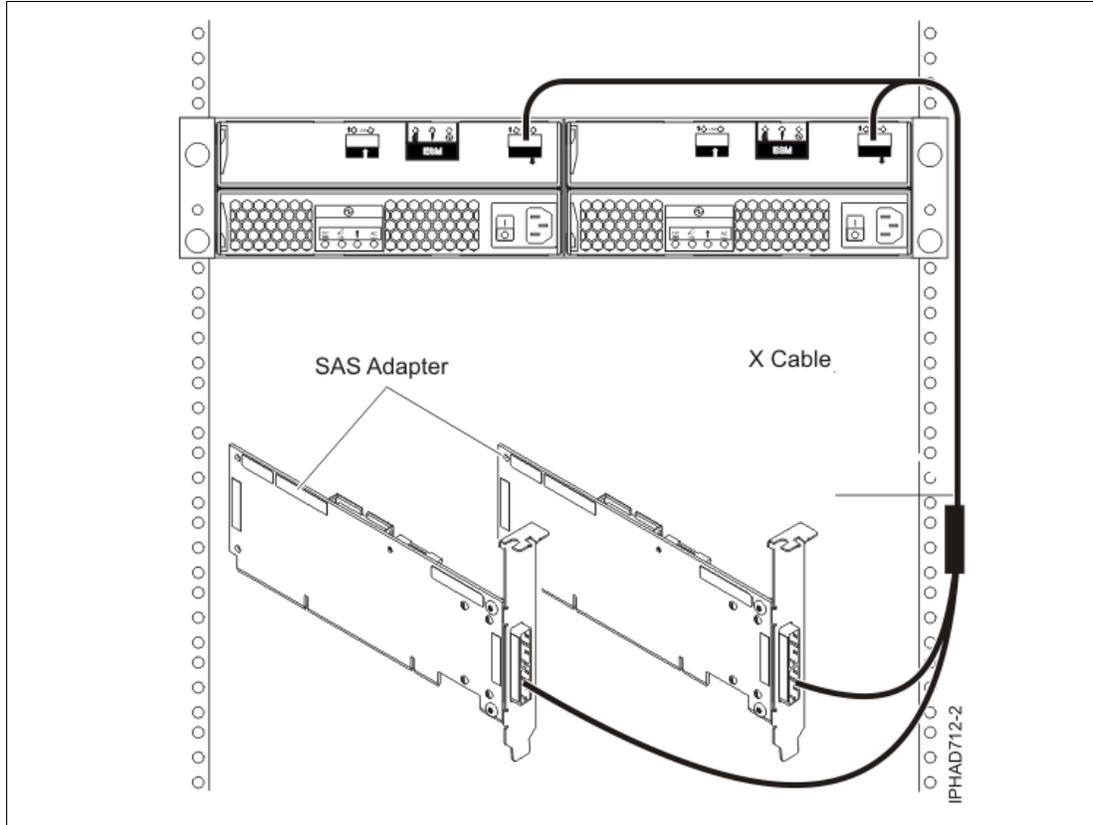
Figure 2-23   Maximum performance example

Attached to the new SAS Controller #5904/#5906/#5908

#5904, #5906 and #5908 are the same cards with different feature numbers, indicating the use in specific expansions. #5906 signifies BSC (blind swap cassette) used in 24-inch expansion and #5908 signifies BSC used in 19-inch expansions.

Figure 2-24 on page 106 provides the schematic to connect the controller to the EXP12S #5886 expansion.

Key features for the card and its connectivity:

- Huge Cache SAS RAID Disk/SSD Controller. 1.5 GB write cache and 1.6 GB read cache.
- Critical to quickly moving IBM i customers off use of SCSI disk drives. Provides great disk performance (also true for AIX/Linux).
- Replacement for IBM i SCSI 1.5 GB write cache controller.
- Maximum of 60 HDD (5 #5886 drawers) per adapter on POWER using EE cables.
- POWER5 does not support use of EE cables (max 36 HDD).
- Maximum of 8 SSD per controller.
- "Pairing" of 1.5 GB adapters like the #5902/03 adapters with X cables is not announced. Use mirrored controllers and drives instead for this level of redundancy.
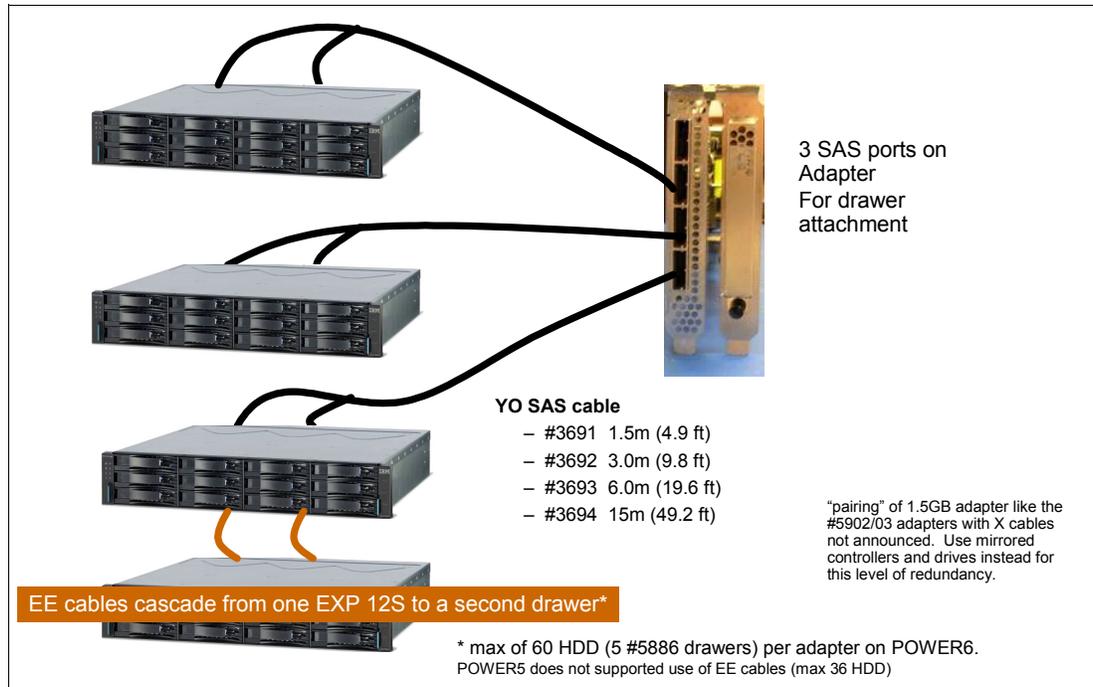
Figure 2-24   Schematic for 5904 connectivity to EXP12S

2.12.5 Multi-initiator and high availability

The terms multi-initiator and high availability (HA) refer to connecting multiple controllers (typically two controllers) to a common set of disk expansion drawers for the purpose of increasing availability. This is commonly done in either of the following configurations:

► HA two system configuration
► HA single system configuration

HA two system configuration

An HA two system configuration provides a high-availability environment for system storage by enabling two systems or partitions to have access to the same set of disks and disk arrays. This feature is typically used with the IBM High-Availability Cluster Multiprocessing application (HACMP™). The IBM HACMP software provides a commercial computing environment that ensures that mission-critical applications can recover quickly from hardware and software failures.

> **Note:** IBM HACMP has some offerings under the HACMP support. During 2008 HACMP has been rebranded as:
>
> ► PowerHA for AIX
> ► PowerHA XD for AIX
> ► PowerHA for Linux
>
> See the following Web site for more information about these product offerings:
>
> http://www-03.ibm.com/systems/power/software/availability/

The HA two system configuration is optimized for using disk arrays. The disks must be formatted to 528 bytes per sector. Any RAID level, or combination of RAID levels, can be used.

Use of disks without RAID (referred to as just basic old disks (JBOD)) is also possible. The disks must be formatted to 512 bytes per sector. This JBOD alternative is supported only on particular controllers and requires unique setup and cabling.

2.12.6 HA single system configuration

An HA single system configuration provides for redundant controllers from a single system to the same set of disks and disk arrays. This feature is typically used with the Multi-Path I/O (MPIO). MPIO support is part of AIX and can be used to provide a redundant IBM SAS RAID Controller configuration with RAID protected disks.

When using an HA single system configuration, the disks must be formatted to 528 bytes per sector and used in one or more disk array. Any RAID level, or combination of RAID levels, can be used. Disks formatted to 512 bytes per sector are not supported in an HA single system configuration.

Table 2-27 summarizes possible configurations.

Table 2-27   Possible configurations

Multi-initiator configuration	HA two system (for example HACMP)	HA single system (for example MPIO)
RAID (disks formatted 528 bytes per sector)	▶ Maximum of two controllers. ▶ Both controllers must have the same write cache capability and write cache sizes. ▶ Both controllers must support HA Two System RAID. ▶ Controllers are in different systems or partitions.	▶ Maximum of two controllers. ▶ Both controllers must have the same write cache capability and write cache sizes. ▶ Both controllers must support HA single system RAID. ▶ Controllers are in the same system or partition.
JBOD (disks formatted 512 bytes per sector)	▶ Maximum of two controllers. ▶ Both controllers must support HA two system JBOD. ▶ Controllers are in different systems or partitions. ▶ Requires unique setup and cabling.	▶ Not supported.

The following sections describe and illustrate an example of each configuration.

HA one/two system RAID

The two controllers can be different systems or partitions for the two system RAID configuration, or the two controllers can be in the system or partition for the one system RAID configuration, as shown in the Figure 2-25.
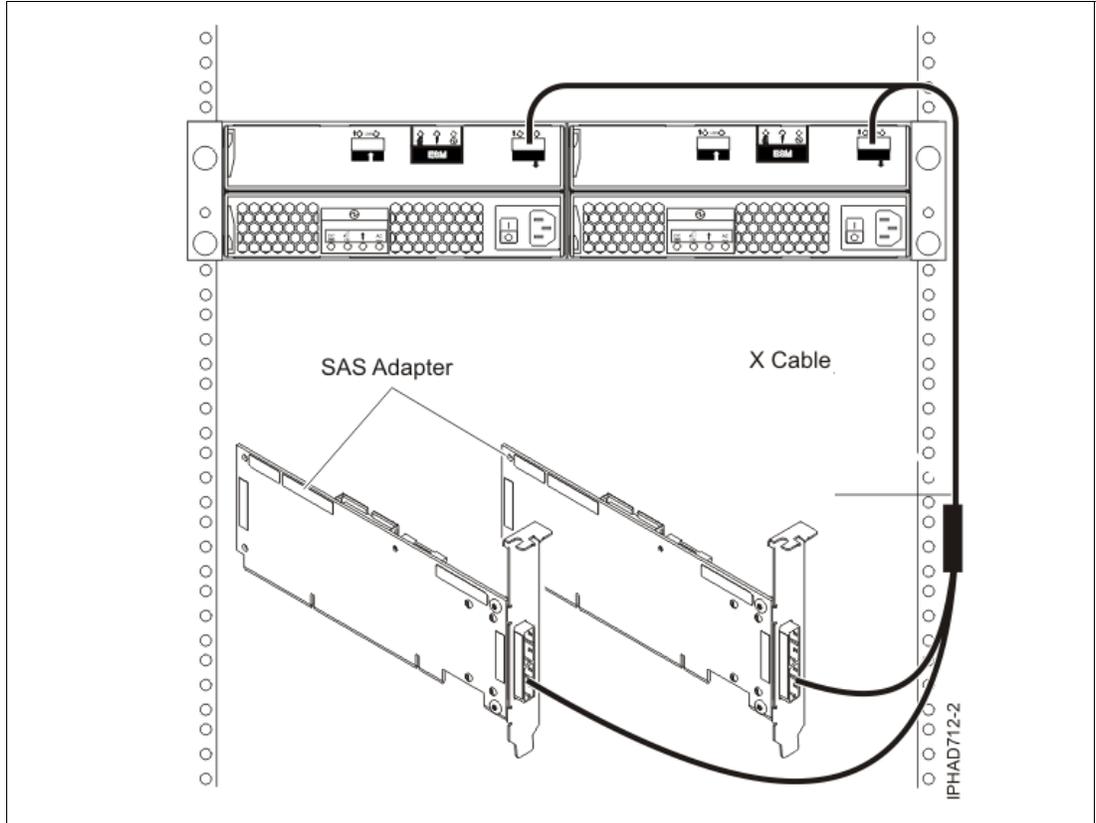

Figure 2-25   HA one/two system RAID

HA two system JBOD

The two controllers have to be in two different systems or partitions for the two system JBOD configuration as shown in Figure 2-26.
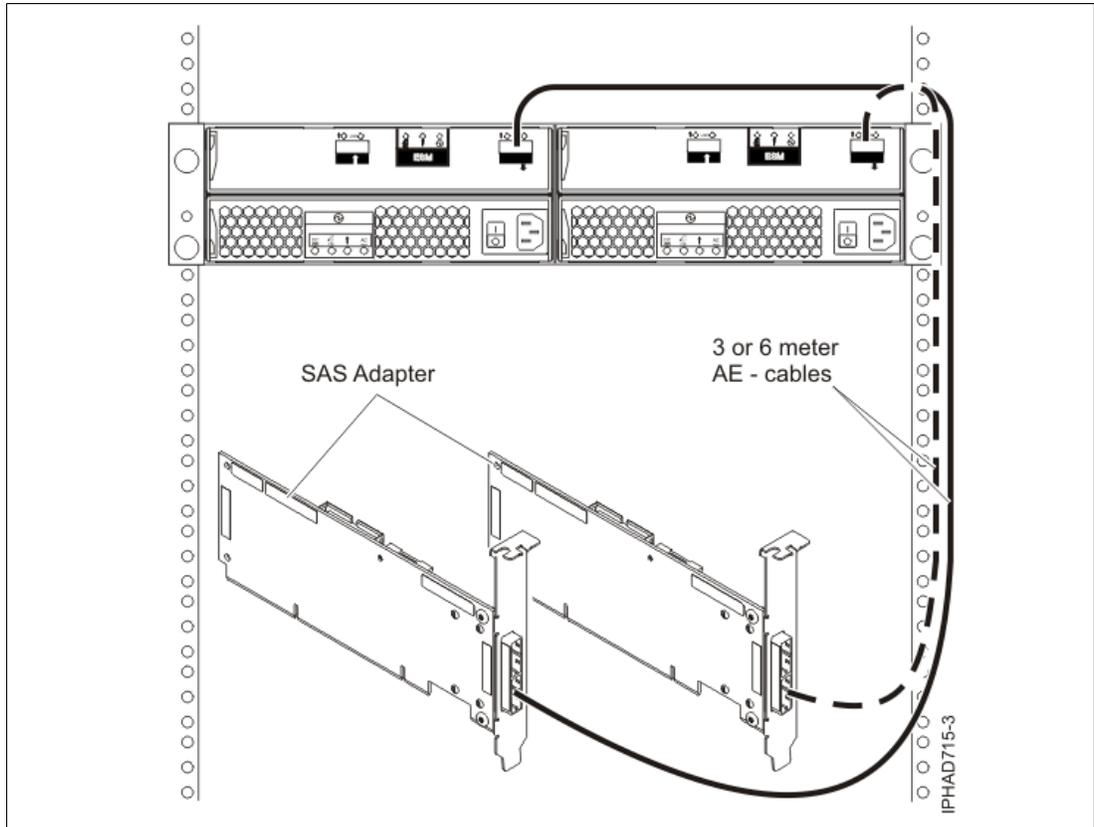
Figure 2-26   Schematic for HA configuration on EXP12S

Note:

► Use of the Multi-initiator and HA functionality requires controller and AIX software support. Specific controllers are intended only to be used in either an HA two system RAID or HA single system RAID configuration.
► Controllers connected in a RAID configuration must have the same write cache size (given they support write cache). A configuration error will be logged if the controllers' write caches are not the same size.

2.13 5802 PCIe 12X I/O drawer

5802 12X I/O drawers expand server I/O performance and capability by supporting higher-speed connections to Power 520 and Power 550 servers and by supporting PCIe 8x adapters and SFF disks.

The 5802 12X I/O drawer is an expansion drawer with 18 SAS SFF Storage Slots as well as 10 PCIe adapter slots. The SAS disks are accessible from the front of the 5802. This I/O drawer takes up a 4 EIA units (4U) space in a 19 inch rack.

This drawer attaches to the central electronics complex via 12X DDR cables (#1862, #1864, or #1865). When SFF disks are installed, they are driven by at least one SAS PCIe adapter and SAS AT cable (#3688). Using a mode switch, the 18 SFF slots can be configured as one group of 18 slots (AIX/Linux), or two groups or 9 slots (AIX/IBM i/Linux), or four groups of 4 or 5 slots (AIX/Linux). 5802 12X I/O drawer characteristics include:

► Modular SAS disk expansion drawer
► Up to eighteen 2.5 inch SFF SAS disk drives
► Support 10 000 rpm disks for AIX, Linux, VIOS
► Support 15 000 rpm disks for AIX, Linux, VIOS and IBM i
► Redundant hot-plug power and cooling with dual line cords
► Up to 2 times faster than PCI-X DDR
► Maximum 2 drawers per loop

2.13.1 5802 PCIe 12X I/O drawer physical description

Table 2-22 lists the 5802 PCIe 12X I/O Drawer attributes.

Table 2-28   #5802 PCIe 12X I/O Drawer Attributes

Dimension	One CEC drawer
Height	174 mm (6.85 in.)
Width	444.5 mm (17.5 in.)
Depth (including front bezel)	711.2 mm (28 in.)
Weight	54 kg (120 lb)

2.13.2 5802 PCIe 12X I/O drawer operating environment

Table 2-29 lists the general system specifications of the system unit.

Table 2-29   5802 PCIe 12X I/O drawer specifications

Description	Range (operating)
Operating temperature range	10° to 38° C (50 to 104 F)
Operating temperature recommended	20° to 28° C (68 to 82 F)
Relative humidity	20% to 80% (allowable) 34% to 54% (recommended)
Noise level	52 bels idle / 52 bels operating
Operating voltage	100-127 V or 200-240 V AC at 50-60 Hz Single Phase

Description	Range (operating)
Maximum power consumption	745 watts (maximum)
Maximum power source loading	0.768 kVa (maximum)
Maximum thermal output	2542 BTU[a]/hr (maximum)
Maximum altitude	3048 m (10 000ft)

a. British Thermal Unit (BTU)

2.13.3 #5802 front and rear views

Figure 2-27 shows the front view of the 5802 PCIe 12X I/O drawer.
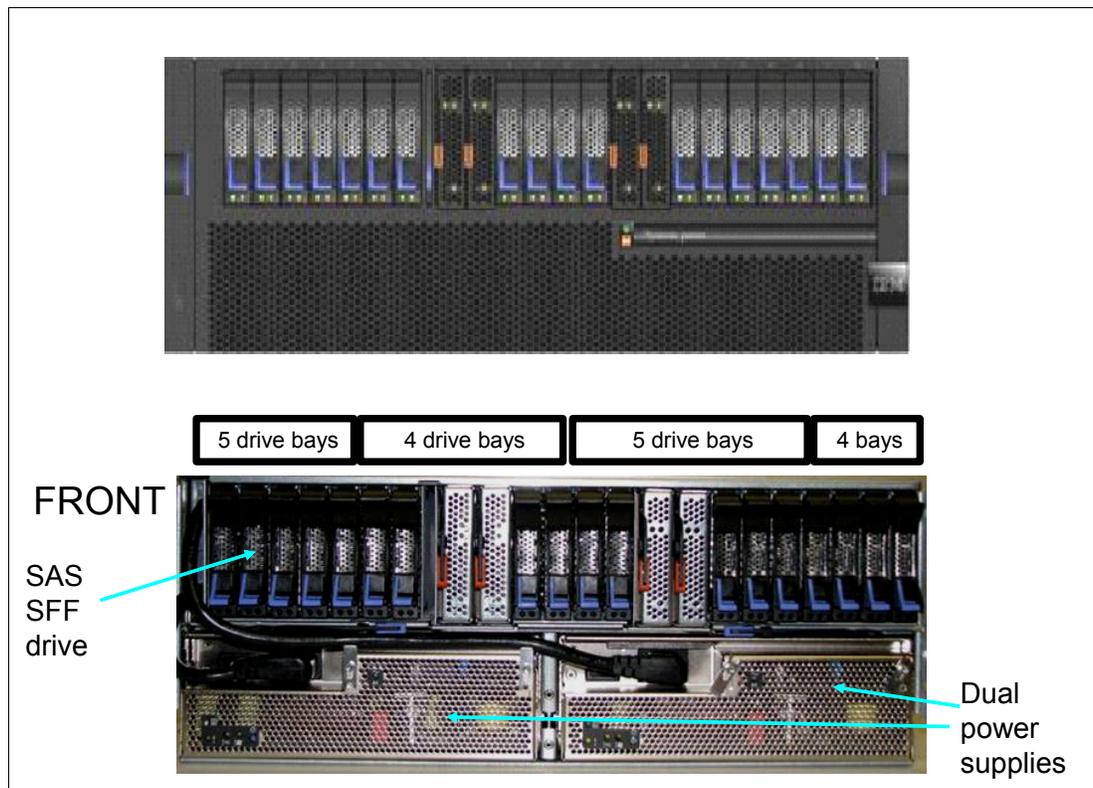
Figure 2-27   Front view #5802 PCIe 12X I/O Drawer

Figure 2-28 shows the rear view of the 5802 PCIe 12X I/O drawer.
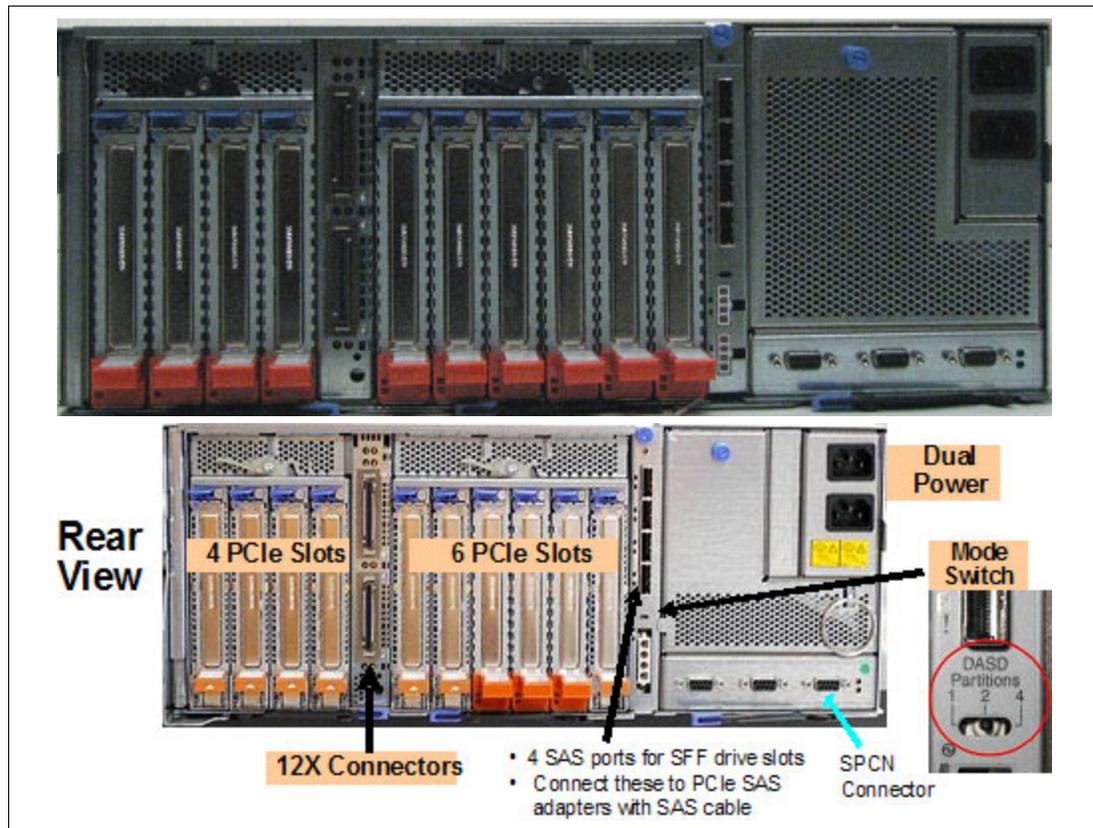
Figure 2-28   Rear view #5802 PCIe 12X I/O drawer

2.13.4 Dividing SFF drive bays in PCIe 12X I/O drawer

Disk drive bays in a PCIe 12X I/O drawer can be configured as one, two, or four set. This allows for partitioning of disk bays. Disk bay partitioning configuration can be done via physical mode switch on the PCIe 12X I/O drawer.

Note: Mode change using a physical mode switch requires power-off/on of the drawer.

Figure 2-19 on page 99 indicates the mode switch in the rear view of the #5802 PCIe 12X I/O drawer.

Each disk bay set can be attached to its own controller or adapter. #5802 PCIe 12X I/O drawer has four SAS connections to drive bays. It connects to PCIe SAS adapters or controllers on the host system.

Figure 2-29 on page 113 shows the configuration rule of disk bay partitioning in #5802 PCIe 12X I/O drawer. There is no specific feature code for mode switch setting.

Note: The IBM System Planning Tool supports disk bay partitioning. Also, IBM configuration tool accepts this configuration from IBM System Planning Tool and passes it through IBM manufacturing using Customer Specified Placement (CSP) option.
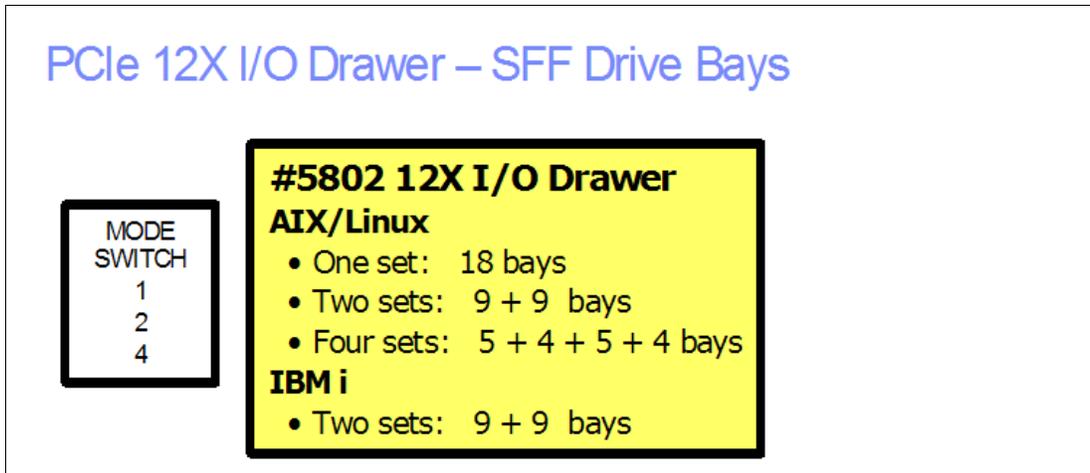
Figure 2-29   Disk bay partitioning in #5802 PCIe 12X I/O drawer

2.13.5 Host system specific information

The following sections provide more details about the PCIe 12X I/O drawer support according to the machine types that are determined by the ordered primary operating system.

> **Note**: New disk drives are announced continuously. In addition, older disk drives are removed from marketing. You should periodically review recent announcement letters to determine the currently marketed and supported disk devices, in all supported I/O enclosures.

Attached to 8203-E4A and 8204-E8A

The PCIe 12X I/O drawer must be mounted in a 19 inch rack, such as the IBM 7014-T00 or 7014-T42. To connect a PCIe 12X I/O drawer to an 8203-E4A or 8204-E8A system unit, a 12X GX connection adapter is needed.

► #5609 Dual ports 12X+ GX loop adapter: This GX loop adapter can support double data rate (DDR) of initial #5616 12X loop adapter. This adapter is available for four or more cores configuration of Power 520 and Power 550 only.
► #5616 Dual ports 12X GX loop adapter: This GX loop adapter can support single data rate (SDR). This adapter is available for any Power 520 and Power 550 configuration except one core Power 520 configuration.

Figure 2-30 on page 114 and Figure 2-31 on page 114 shows the 12X loop cabling with #5802 PCIe 12X I/O Drawers.

A PCIe 12X I/O drawer cannot mix with a #5796 or a 7413-G30 PCI-X 12X I/O drawer on the same loop. Compared with a #5802 PCIe 12X I/O drawer, #5796 and 7413-G30 PCI-X 12X I/O drawers can be maximum four drawers on a loop. But the #5802 PCIe 12X I/O drawer has more PCI slots per drawer. #5796 and 7413-G30 PCI-X 12X I/O drawers have only six PCI slots per drawer.

Figure 2-30   Two #5802 PCIe 12X I/O Drawers on a loop

Figure 2-31   PCIe 12X I/O drawer loop configurations with Power 520 and Power 550

Depending on the required configuration a new and different set of 12X DDR cables is needed to connect the PCIe 12X I/O drawer to the system or another PCIe 12X I/O drawer. A #5802 PCIe 12X I/O drawer is not supported by earlier 12X SDR cables, but a PCI-X 12X I/O drawer is supported by new 12X DDR cables. A list of cables is provided in Table 2-30 on page 115.

Table 2-30   New 12X DDR connection cables

Feature	Description
1861	0.6m 12X DDR cable for PCIe and PCI-X 12X I/O drawer
1862	1.5m 12X DDR cable for PCIe and PCI-X 12X I/O drawer
1863	2.5m 12X DDR cable for PCIe and PCI-X 12X I/O drawer
1864	8.0m 12X DDR cable for PCIe and PCI-X 12X I/O drawer
1865	3.0m 12X DDR cable for PCIe and PCI-X 12X I/O drawer

You can mix old and new cables on 12X loop with older PCI-X 12X I/O drawers
(#5796/5797/5798) at your location, but 4.7 GHz Power 520 and 5.0 GHz Power 550 do not
support this. IBM configuration tools also do not support this in order to avoid mixing cables
on the same MES order, since IBM manufacturing could not know which cable goes where.

Note: #1863 2.5 m 12X DDR cable is not supported by #5802 PCIe 12X I/O drawer. This
cable is for 24 inch PCIe 12X I/O drawers (#5803/5873).

New 12X DDR cables have the same dark green color connector as old 12X SDR cable, but
have different labels. Installation of cables should be done carefully; physically the old cables
can be connected to the new PCIe 12X I/O drawers, but they will not work. The size and
keying is identical.

Figure 2-32 shows the new 12X DDR cable has label "D" indicating DDR connection.
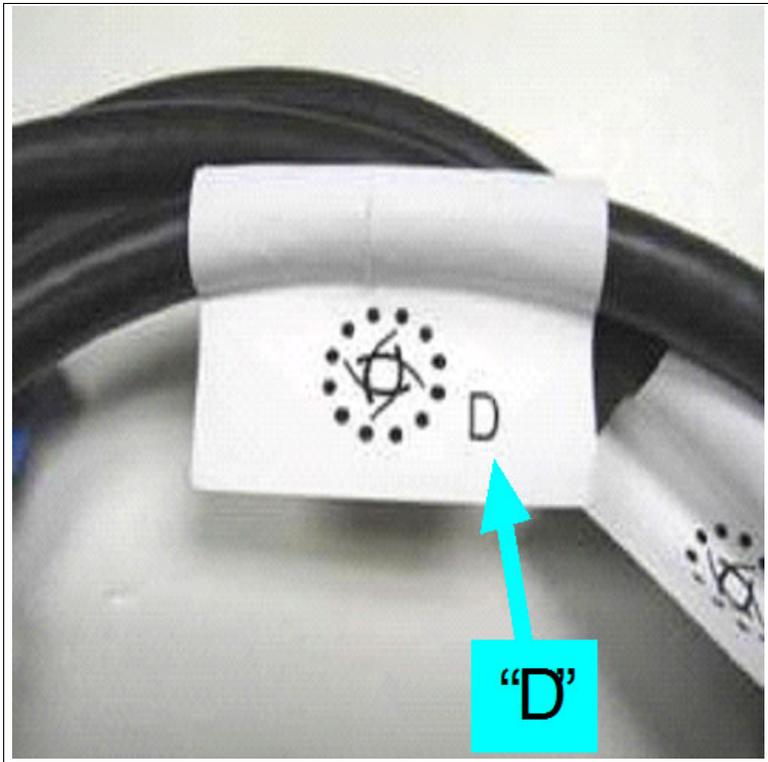
Figure 2-32   New 12X DDR connection cable

To use SFF SAS disks in a PCIe 12X I/O drawer with an 8203-E4A or 8204-E8A system unit, an additional SAS RAID adapter is needed. Table 2-31 provides the current list of available adapters.

Table 2-31   SAS adapters

Feature	Description
5901[1]	PCIe DDR Dual - x4 3 Gb SAS RAID Adapter supports RAID 0 and 10
5903[2]	PCIe DDR Dual - x4 3 Gb SAS RAID Adapter with 380MB Cache supports RAID 0, 5, 6 and 10.

1. One #5901 is needed per group of SFF disk drive bay. With AIX/Linux, up to four #5901 accessing four sets of SFF drive bays. With IBM i, only two (two groups of drive bays) are supported. #5901 has two ports. 2nd port can drive a second group of drives in the #5802 PCIe 12X I/O Drawer or it can drive a #5886 EXP12S Disk drawer (or two EXP12S if cascaded using EE cables).
2. Pair of #5903 provides redundancy. Adapters are linked together via connections in #5802 PCIe 12X I/O Drawer and write cache contents are mirrored. If one adapter fails or disconnected, contents of write cache are written out and write cache is disabled until pairing restored.   Usually significant performance impact is possible without write cache, especially if running RAID-5 or RAID-6. Can also use the 2nd pair of SAS ports on the pair of #5903 to attach to #5886 EXP 12S.

For redundancy purposes, a pair of #5901 PCIe SAS RAID adapters can be used, but this configuration does not support four sets of SFF disk bay partitions.

Figure 2-33 on page 117 shows a #5903 PCIe SAS RAID adapter with 380 MB Cache and cabling to SAS ports on a #5802 PCIe 12X I/O drawer.

Whenever #5901/#5903 SAS RAID adapter is installed on #5802 PCIe 12X I/O drawer, #3688 SAS AT cable is needed. This #3688 SAS AT cable is needed per port of the SAS adapter.

#5901/#5903 SAS RAID adapter card is always in the same drawer as the disk drives.

Paired SAS Disk Adapters & #5802 PCIe 12X I/O Drawer

Pair of #5903 PCIe 380 MB cache

#5903 SAS adapter

#5903 SAS adapter

SAS AE cables #3688
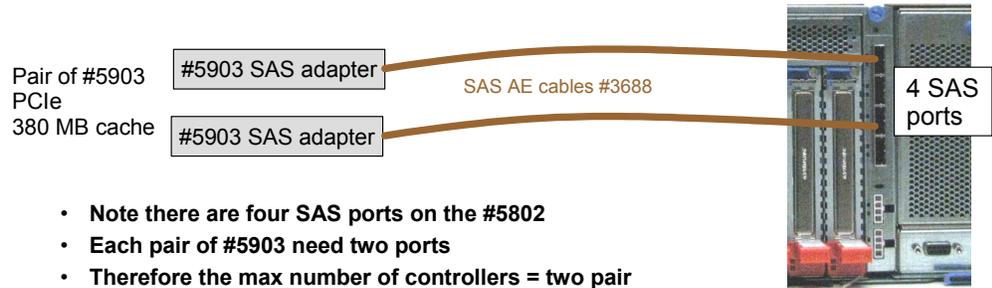
4 SAS ports

- Note there are four SAS ports on the #5802
- Each pair of #5903 need two ports
- Therefore the max number of controllers = two pair
- Thus the max number of groups/partitions per #5802 = two using #5903
- If want 4 groups of disk slots, use four unpaired #5901

Pair of #5901 PCIe 0 MB cache

SAS adapter

SAS adapter

- Optionally, can pair the #5901 PCIe SAS Adapters controller for redundancy
- "Optional" pairing because no write cache to protect
- If paired, same two partition maximum as with #5903 adapter

Figure 2-33   SAS Disk Adapters and #5802 PCIe 12X I/O Drawer

Figure 2-34 on page 118 illustrates #5901/#5903 PCIe SAS RAID adapter cabling rules with SAS ports in #5802 PCIe 12X I/O drawer. If both of the dual ports in the same #5901/#5903 PCIe SAS RAID adapter are used, do not connect adjacent ports in SAS ports in #5802 PCIe 12X I/O drawer. The SAS RAID adapter slot used is not important. However, plug rules state use slot number 1, which is the left-most slot in #5802 PCIe 12X I/O drawer from rear view, first.
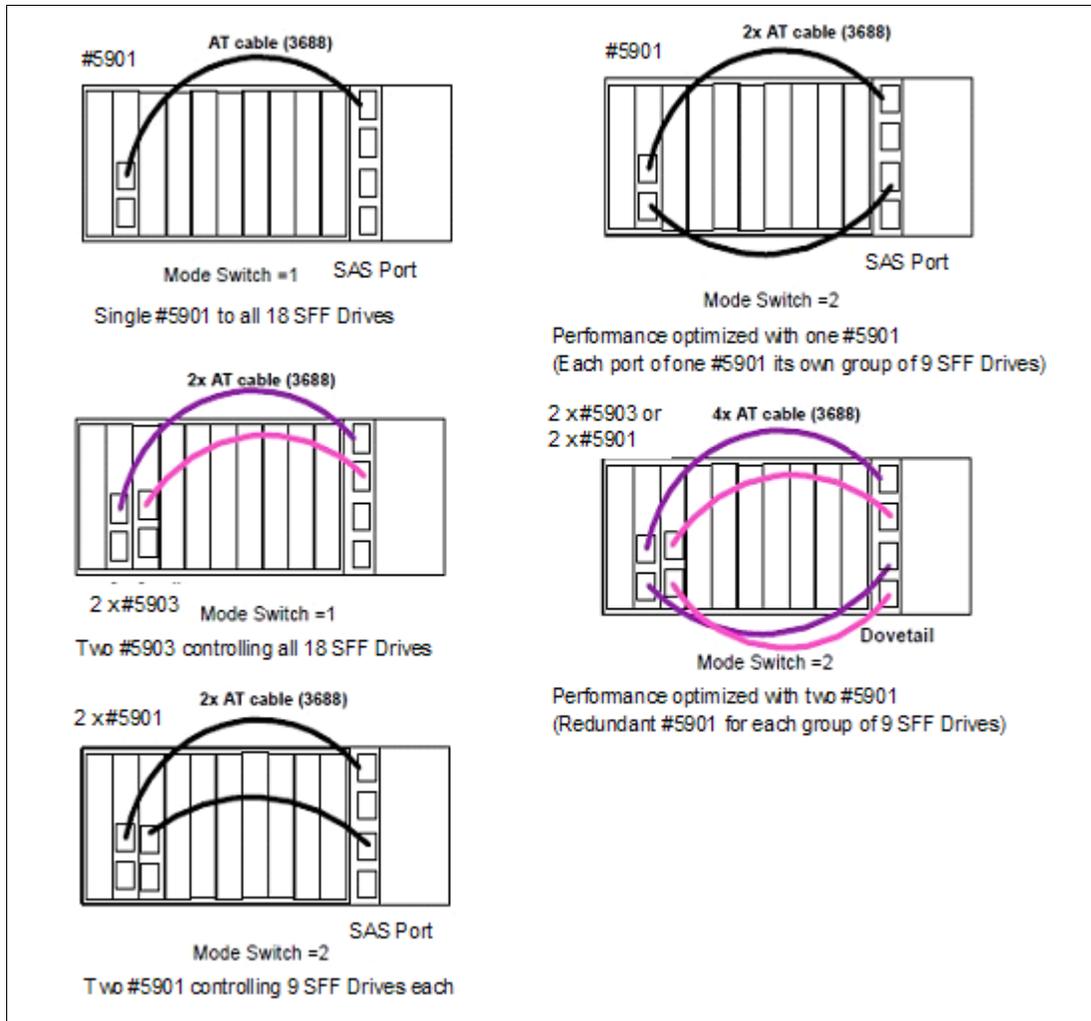
Figure 2-34   #5901/#5903 PCIe SAS RAID adapter plug and cabling rules in #5802 PCIe 12X I/O drawer

> **Note:** Dual #5802 PCIe 12X I/O drawers for I/O planer fault tolerance is not supported.

Load source consideration

Feature #0726 specifies that the Load/Source DASD are placed within the #5802 PCIe 12X I/O Drawer. There is no specific load source consideration and the load source can be put in any of the SFF SAS disk drive bays.

2.14 Hardware Management Console

The Hardware Management Console (HMC) is a dedicated workstation that provides a graphical user interface for configuring, operating, and performing basic system tasks for the POWER6 processor-based systems (as well as the POWER5 and POWER5+ processor-based systems) that function in either non-partitioned, partitioned, or clustered environments. In addition, the HMC is used to configure and manage partitions. One HMC is capable of controlling POWER5, POWER5+, and POWER6 processor-based systems.

At the time of writing, one HMC supports up to 48 POWER5, POWER5+, and POWER6 processor-based systems and up to 254 LPARs using the HMC machine code Version 7.3. For updates of the machine code and HMC functions and hardware prerequisites, refer to the following Web site:

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

POWER5, POWER5+, and POWER6 processor-based system HMCs require Ethernet connectivity between the HMC and the server's service processor. Furthermore, if dynamic LPAR operations are required, all AIX 5L, AIX V6, and Linux partitions must be enabled to communicate over the network to the HMC. Ensure that at least two Ethernet ports are available to enable public and private networks:

► The HMC 7042 Model C06 is a deskside model with one integrated 10/100/1000 Mbps Ethernet port, and two additional PCI slots.

► The HMC 7042 Model CR4 is a 1U, 19 inch rack-mountable drawer that has two native 10/100/1000 Mbps Ethernet ports and two additional PCI slots.

> **Note:** The IBM 2-Port 10/100/1000 Base-TX Ethernet PCI-X Adapter (#5706) should be ordered to provide additional physical Ethernet connections.

For any logical partition in a server, it is possible to use a Shared Ethernet Adapter set in Virtual I/O Server or Logical Ports of the Integrated Virtual Ethernet card, for a unique or limited number of connections from the HMC to partitions. Therefore, a partition does not require its own physical adapter to communicate to an HMC.

It is a good practice to connect the HMC to the first HMC port on the server, which is labeled as HMC Port 1, although other network configurations are possible. You can attach a second HMC to HMC Port 2 of the server for redundancy (or vice versa). Figure 2-35 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 the *Hardware Management Console V7 Handbook*, SG24-7491.
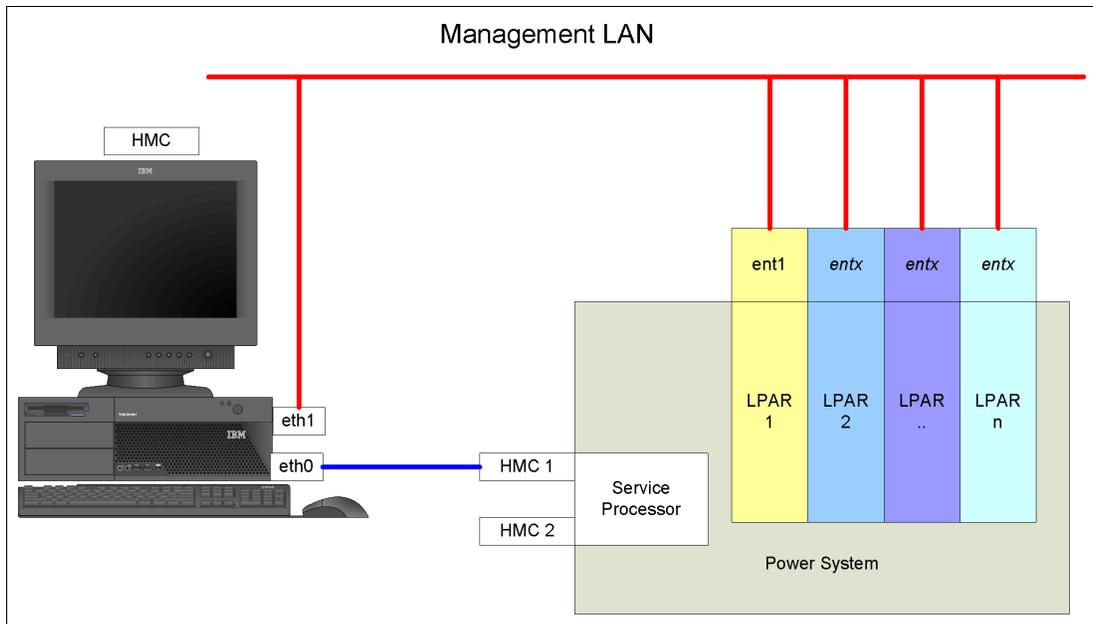

Figure 2-35   HMC to service processor and LPARs network connection

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 the service processor of the managed server does not receive DHCP reply before time-out, predefined IP addresses will set up on both ports. Static IP address allocation is also an option. You can configure the IP address of the service processor ports with a static IP address by using the Advanced System Management Interface (ASMI) menus.

> **Note:** The 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 default IP addresses:
>
> – Service processor Eth0 or HMC1 port is configured as 169.254.2.147 with netmask 255.255.255.0
>
> – Service processor Eth1 or HMC2 port is configured as 169.254.3.147 with netmask 255.255.255.0
>
> More information about the Service Processor can be found in 4.5.1, "Service processor" on page 182.

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

▶ Generating or importing System Plans

The HMC provides both graphical and command line interface for all management tasks. Remote connection to the HMC using a Web browser (as of HMC Version 7, previous versions required a special client program, called WebSM) or SSH are possible. 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.

2.14.1 High availability using the HMC

The HMC is an important hardware component. IBM PowerHA high availability cluster software can be used to execute dynamic logical partitioning operations or activate additional resources (where available), thus becoming an integral part of the cluster.

If redundant HMC function is desired, the servers can be attached to two separate HMCs to address availability requirements. All HMCs must have the same level of Hardware Management Console Licensed Machine Code Version 7 (#0962) to manage POWER6

processor-based servers or an environment with a mixture of POWER5, POWER5+, and POWER6 processor-based servers. The HMCs provide a locking mechanism so that only one HMC at a time has write access to the service processor. Depending on your environment, you have multiple options to configure the network. Figure 2-36 shows one possible highly available configuration.
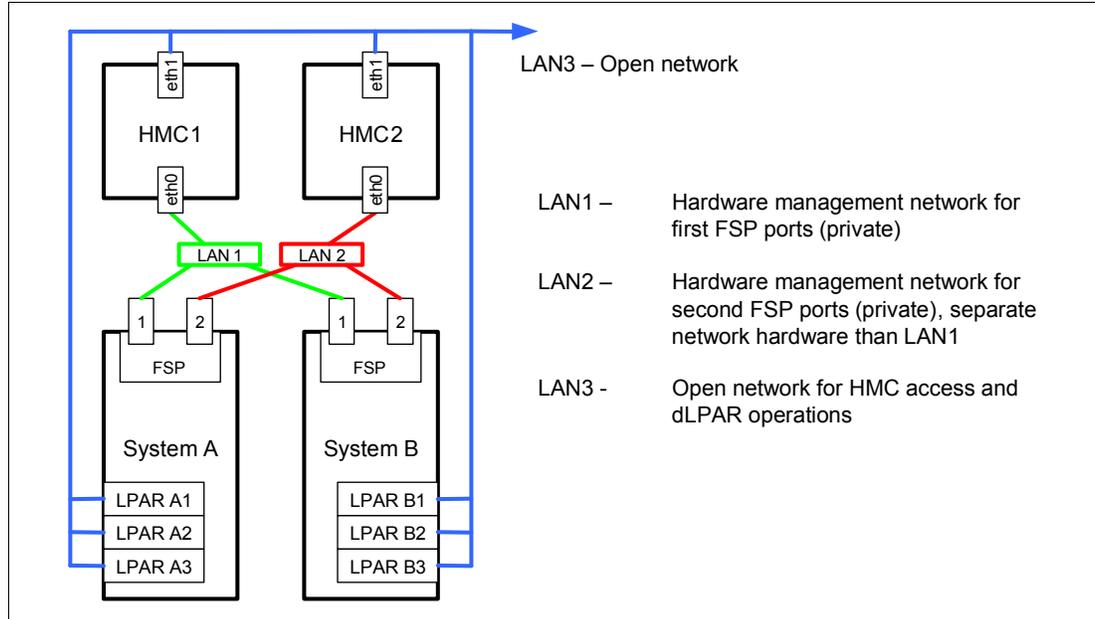

Figure 2-36   Highly available HMC and network architecture

Note that only hardware management networks (LAN1 and LAN2) are highly available in this figure; this was done to keep the diagram simple. However, management network (LAN3) can be made highly available by using a similar concept and adding more Ethernet adapters to LPARs and HMCs.

2.15 Service information

Power 520s are client setup servers (CSU) and are shipped with materials to assist in the general installation of the server. The server cover has a quick reference service information label that provides graphics to 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 to install it into the system, to open or to 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 have 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 server.
Blue and terra-cotta	Terra-cotta takes precedence over this color combination, and the rules for a terra-cotta-only touch point apply.

> **Important:** Adhere to the touch point colors on the system! Not doing so can compromise your safety and damage the system.

2.15.2 Control (operator) panel

The service processor provides an interface to the control panel (also referred to as the operator panel or simply the *Ops* panel in various sets of documentation) that displays server status and can be used to perform manual diagnostic activities and system start and stop operations without an HMC. Figure 2-37 on page 123 shows the control panel with blue tab (A) pushed to disengage the panel from its normal "display only" position and pivoted into place for use. To use the control panel you must disengage it from its normal display only resting position within the system unit.

The lower portion of the figure identifies the buttons of the panel ready to be used for manual functions.
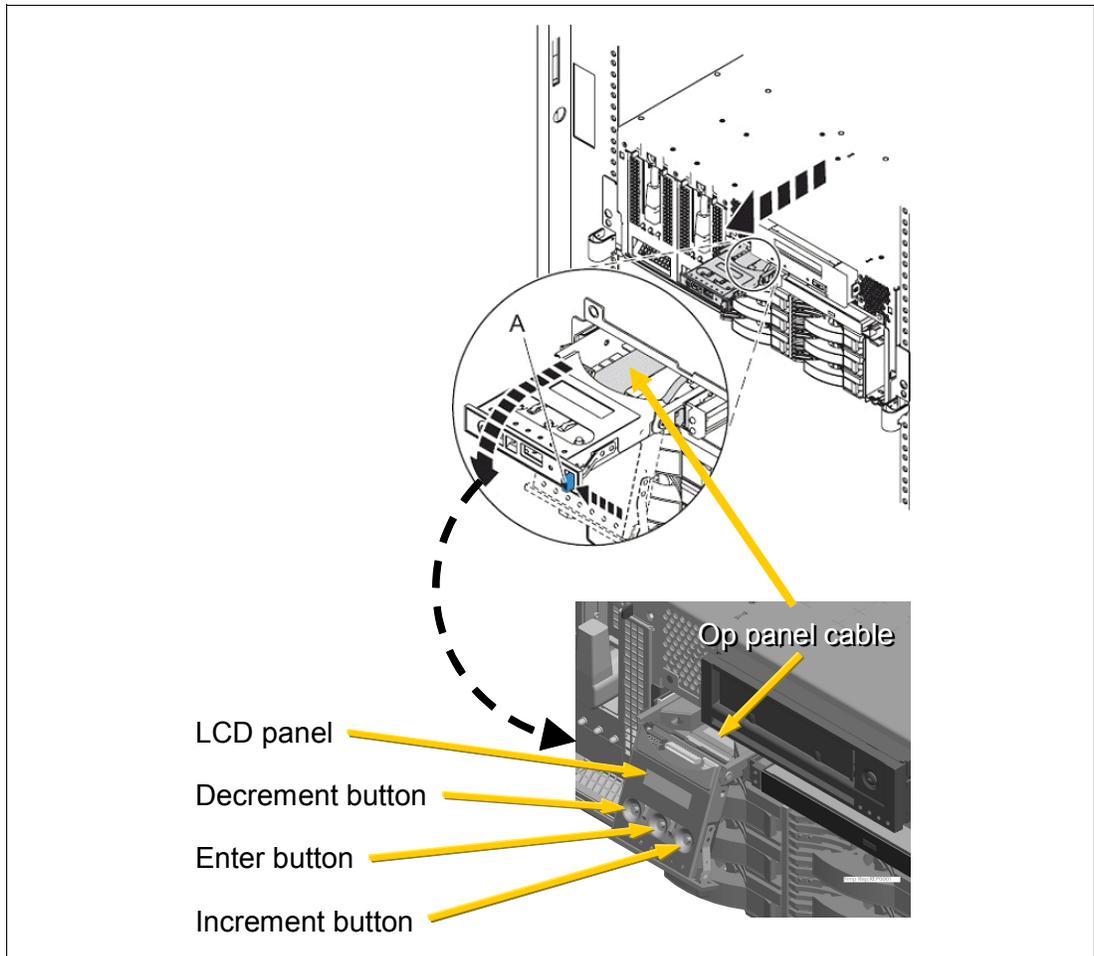
*Figure 2-37   Operator control panel physical details*

Figure 2-38 on page 124 represents a view of the control panel. The upper part of the figure shows the normal operation position. The lower part of the figure shows the panel disengaged from it normal position and ready for manual operation.

Note: For servers managed by the HMC, you would normally used the HMC interface to perform control panel functions.
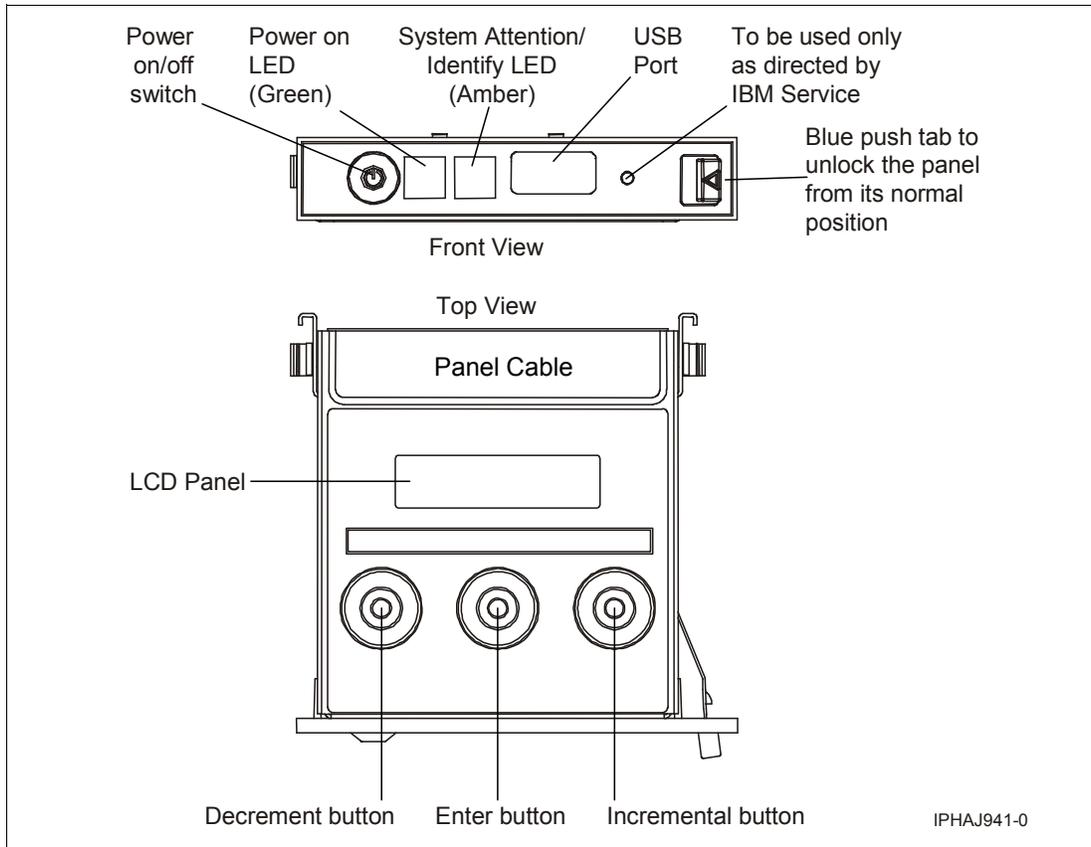
Figure 2-38   Idealized view of control panel: normal position, manual use position views

Manual mode usage is described in the next topic.

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 the selected IPL type, system operating mode, and IPL speed

► Function 02: Select the 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 client service switch 1 (Function 25), followed by service switch 2 (Function 26).

Function 30 – System unit (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. Table 2-32 shows an example of how to use Function 30.

Table 2-32   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 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 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.

2.15.3 Securing a rack-mounted system in a rack

The *optional* rack-mount drawer rail kit is a unique kit designed for use with the Power 520 rack-mounted model. No tools are required to install the Power 520, 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 the unit to easily 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.

2.15.4 Servicing 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 Figure 2-39.
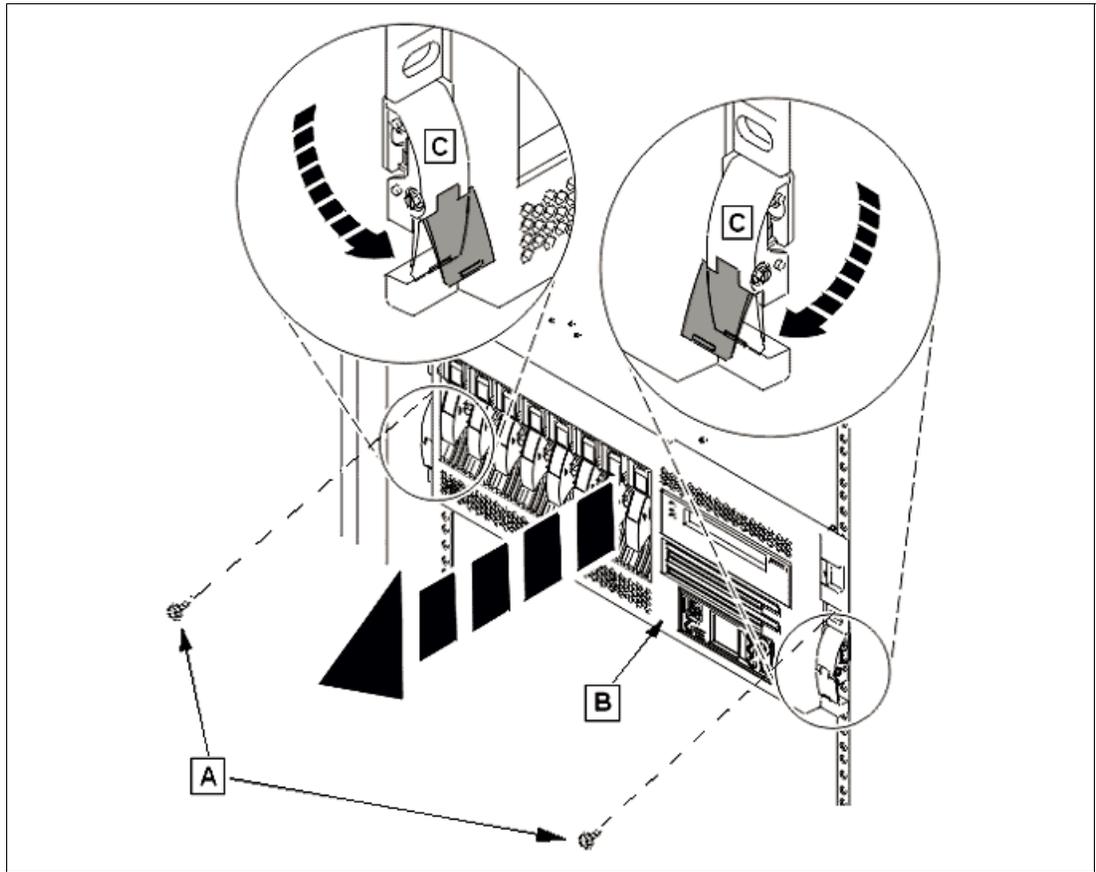
Figure 2-39   Pull the Power 520 to the service position

3. Release the rack latches (C) on both the left and right sides, as shown in the figure.

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 31.75 kg (70 lb.). Ensure that you can safely support this weight when removing the server unit from the system rack.

Refer to the IBM Systems Hardware Information Center for more information or to view available video clips that demonstrate several of the maintenance and repair procedures.

2.15.5 Cable-management arm

The Power 520 rack-mounted models are shipped with a cable-management arm to route all the cables through the hooks along the cable arm and to secure them with the straps that are provided. The cable-management arm simplifies the cable management in case of a service action that requires you to pull the rack-mounted system from the rack.

2.16 System firmware

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

> **Note:** Normally, installing the server firmware fixes through the operating system is a non concurrent process.

Temporary and permanent firmware sides

The service processor maintains two copies of the server firmware:

- ► One copy is considered the permanent or backup copy and is stored on the permanent side, sometimes referred to as the *p* side.

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

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

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

The following examples are the output of the `lsmcode` command for AIX and Linux, showing the firmware levels as they are displayed in the outputs.

- ► AIX:

  ```
  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)
  ```

IBM i 6.1 includes the command Display Firmware Status (DSPFMWSTS). An example of its output is shown in Figure 2-40 on page 128.

```
                       Display Firmware Status

Service partition . . . . . . . . . . . . . :     Yes
Firmware update policy  . . . . . . . . . . :     *HMC
Server IPL source . . . . . . . . . . . . . :     Temporary
Firmware product ID/release . . . . . . . . :     5733905  V1R1M0



            --Server firmware--  -Operating system--
            Fix        PTF       Fix        PTF        Update
Copy        pack       ID        pack       ID         Status
*ACTIVE     SF240_320  *UNKNOWN
*TEMP       SF240_320  *UNKNOWN  *NONE      *NONE      None
*PERM       SF240_320  *UNKNOWN  *NONE      *NONE      None




                                                              Bottom
```

Figure 2-40   IBM i 6.1 Display Firmware level example

Firmware level EL340_039 or later is required for support of 4Q 2008 hardware announcements, including 8203-E4A support of IBM i.

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 it. 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 for POWER5 and POWER6 servers, start at:

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

This is a common Microcode download site. You can select the server, for example, 8203-E4A, 8204-E8A, 9407-M15, 9408-M25, or 9409-M50. Follow the information presented in the next set of windows to review what is available and how to download the appropriate software.

Receive server firmware fixes using an HMC

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

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

► The HMC or server is connected to the Internet.

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

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

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

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

1. Ensure that you have a connection to the service provider (if you have an Internet connection from the HMC or server).

2. Determine the available levels of server and power subsystem firmware.

3. Create the optical media (if you do not have an Internet connection from the HMC or server).

4. Use the Change Internal Code wizard to update your server and power subsystem firmware.

5. Verify that the fix installed successfully.

For a detailed description of each task, refer to the IBM Systems Hardware Information Center Web site at:

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

Follow these links: **Power Systems** → your MTM (for example, **8203-E4A**) → **Troubleshooting, service, support of POWER6-based processors** → **Getting fixes** → **Getting firmware fixes.**

You can then select from several options that meet your operating environment requirements.

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, IBM i, or Linux.

To do this, complete the following tasks:

1. Determine the existing level of server firmware using the `lsmcode` command.

2. Determine the available levels of server firmware.

3. Get the server firmware.

4. Install the server firmware fix to the temporary side.

5. Verify that the server firmware fix installed successfully.

6. Install the server firmware fix permanently (optional).

> **Note:** You can use the HMC or the Advanced Management Interface (AMI) to view firmware levels on your system.
>
> You can also view existing levels of server firmware using the `lsmcode` command under AIX or Linux. You need to have the following service tools installed on your server:
>
> ► AIX
>
>   You must have AIX diagnostics installed on your server to perform this task. AIX diagnostics are installed when you install AIX on your server. However, it is possible to deselect the diagnostics. Therefore, you need to ensure that the online AIX diagnostics are installed before proceeding with this task.
>
> ► Linux
>
>   – Platform Enablement Library: librtas-*nnnnn*.rpm
>   – Service Aids: ppc64-utils-*nnnnn*.rpm
>   – Hardware Inventory: lsvpd-*nnnnn*.rpm
>
>   Where *nnnnn* represents a specific version of the RPM file.
>
>   If you do not have the service tools on your server, you can download them at the following Web site:
>
>   `https://www14.software.ibm.com/webapp/set2/sas/f/lopdiags/home.html`
>
> IBM 6.1 is required to use the Display Firmware Status (DSPFMWSTS) command.

2.16.1 Service processor

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

The service processor is used to monitor and manage the system hardware resources and devices. The service processor offers 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 Power 520 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 169.254.2.147.

  – Service processor Eth1 or HMC2 port is configured as 169.254.3.147.

2.16.2 Hardware management user interfaces

This section provides a brief overview of the Power 520 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 you to change the service processor IP addresses or to apply some security policies and avoid the access from undesired IP addresses or ranges. 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.

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 runtime. However, some of the menu options in the Web interface are unavailable during IPL or runtime 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 runtime.

Accessing the ASMI using an HMC

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

1. Open Systems Management from the navigation pane.
2. From the work pane, select one or more managed systems to work with.
3. From the System Management tasks list, select Operations.
4. From the Operations task list, select Advanced System Management (ASM).

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 (power-on-self-test) indicators (memory, keyboard, network, SCSI, and speaker) that appear across the bottom of the screen. Press the numeric 1 key after the word `keyboard` appears and before the word `speaker` appears.

The SMS menus is useful to defining 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 the case of a network boot, SMS menus are provided to set up the network parameters and network adapter IP address.

HMC

The Hardware Management Console is a system that controls managed systems, including IBM System p5 and IBM System p servers, and logical partitions. To provide flexibility and availability, there are several ways to implement HMCs.

3

Virtualization

As you look for ways to maximize the return on your IT infrastructure investments, consolidating workloads becomes an attractive proposition. IBM Power Systems combined with PowerVM technology are designed to help you consolidate and simplify your IT environment. Virtualization plays a key role in accomplishing this goal.

Virtualization capabilities help you:

- ► Improve server utilization and share I/O resources to reduce total cost of ownership and make better use of IT assets.
- ► Improve business responsiveness and operational speed by dynamically re-allocating resources to applications as needed — to better match changing business needs or handle unexpected changes in demand.
- ► Simplify IT infrastructure management by making workloads independent of hardware resources, thereby enabling you to make business-driven policies to deliver resources based on time, cost, and service-level requirements.

This chapter discusses the virtualization technologies and features on IBM Power Systems:

3.1 POWER Hypervisor

Combined with features designed into the POWER6 processors, the POWER Hypervisor delivers functions that enable other system technologies, including logical partitioning technology, virtualized processors, IEEE VLAN compatible virtual switch, virtual SCSI adapters, and virtual consoles. The POWER Hypervisor is a basic component of the system's firmware and offers the following functions:

► Provides an abstraction between the physical hardware resources and the logical partitions that use them

► Enforces partition integrity by providing a security layer between logical partitions

► Controls the dispatch of virtual processors to physical processors (see 3.2.3, "Processing mode" on page 138)

► Saves and restores all processor state information during a logical processor context switch

► Controls hardware I/O interrupt management facilities for logical partitions

► Provides virtual LAN channels between logical partitions that help to reduce the need for physical Ethernet adapters for inter-partition communication

► Monitors the Service Processor and will perform a reset/reload if it detects the loss of the Service Processor, notifying the operating system if the problem is not corrected

The POWER Hypervisor is always active, regardless of the system configuration and even when an HMC is not connected to the system. It requires memory to support the resource assignment to 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

► Number of physical and virtual I/O devices used by the logical partitions

► Maximum memory values given to the logical partitions

The minimum amount of physical memory to create a partition is the size of the system's Logical Memory Block (LMB). The default LMB size varies according to the amount of memory configured for the system, as shown in Table 3-1.

Table 3-1   Configurable memory and associated default Logical Memory Block size

Configurable memory	Default Logical Memory Block
Less than 4 GB	16 MB
Greater than 4 GB up to 8 GB	32 MB
Greater than 8 GB up to 16 GB	64 MB
Greater than 16 GB up to 32 GB	128 MB
Greater than 32 GB	256 MB

In most cases, the actual requirements and recommendations are between 256 MB and 512 MB for AIX, Red Hat, and Novell SUSE Linux. Physical memory is assigned to partitions in increments of Logical Memory Block (LMB).

The HMC provides an interface to adjust some of the logical memory block sizes on larger configurations. The default values should be changed only by experienced administrators.

The POWER Hypervisor provides the following types of virtual I/O adapters:

► Virtual SCSI

► Virtual Ethernet

► Virtual (TTY) console

Virtual SCSI

The POWER Hypervisor provides a virtual SCSI mechanism for virtualization of storage devices (a special logical partition to install the Virtual I/O Server is required to use this feature, as described in 3.3.2, "Virtual I/O Server" on page 141). 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 PowerVM Edition features.

Virtual Ethernet

The POWER Hypervisor provides a virtual Ethernet switch function that allows partitions on the same server to use 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[1] size and CPU entitlement. Virtual Ethernet support starts with AIX 5L Version 5.3, IBM i 5.4, or the appropriate level of Linux supporting Virtual Ethernet devices (see 3.3.8, "Operating system support for PowerVM" on page 151). 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. The maximum number of VLANs is 4096. Virtual Ethernet adapters can be configured as either untagged or tagged (following the IEEE 802.1Q VLAN standard).

► A partition supports 256 virtual Ethernet adapters. Besides a default port VLAN ID, the number of additional VLAN ID values that can be assigned per Virtual Ethernet adapter is 20, which implies that each Virtual Ethernet adapter can be used to access 21 virtual networks.

► Each partition operating system detects the 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 connectivity outside of the server if a layer-2 bridge to a physical Ethernet adapter is set in one Virtual I/O server partition (see 3.3.2, "Virtual I/O Server" on page 141 for more details about shared Ethernet). Also known as Shared Ethernet Adapter.

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.

---

[1] Maximum transmission unit

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 PowerVM Edition features.

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 that is connected to a system port.

3.2 Logical partitioning

Logical partitions (LPARs) and virtualization increase utilization of system resources and add a new level of configuration possibilities. This section provides details and configuration specifications about this topic.

3.2.1 Dynamic logical partitioning

Logical partitioning (LPAR) was introduced with the POWER4™ processor-based product line and the AIX 5L Version 5.1 operating system on System p models. LPAR was introduced on System i models prior to POWER4 and IBM i (then known as OS/400) V4.5.

This technology originally offered the capability to divide a single physical 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 logical partitioning became available, increasing flexibility by allowing selected system resources, such as processors, memory, and I/O components, to be added and deleted from logical partitions while they are executing. These dynamic capabilities became available on the POWER5 systems running IBM i 5.3 and AIX 5L Version 5.2 as far back as 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.

3.2.2 Micro-Partitioning

Micro-Partitioning technology allows you to allocate fractions of processors to a logical partition. This technology was introduced with POWER5 processor-based systems. A logical partition using fractions of processors is also known as a *Shared Processor Partition* or *Micro-Partition*. Micro-Partitions run over a set of processors called *Shared Processor Pool*. Virtual processors are used to let the operating system manage the fractions of processing power assigned to the logical partition. 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. Note that the meaning of the term *physical processor* in this section is a *processor core*. For example, in a 2-core server there are two physical processors.

When defining a shared processor partition, several options or terms have to be defined:

► The minimum, desired, and maximum processing units. *Processing units* are defined as processing power, or the fraction of time the partition is dispatched on physical processors. Processing units define the capacity entitlement of the partition.

  – A minimum processing capacity value must be available or the partition will not start. Because processing capacity units can be moved to other partitions during operation, the minimum value also permits an active partition to not have processor unit capacity removed to lower than the required value.

  – The desired is the processor capacity the partition normally should be started with.

  – The maximum processor capacity can be used to limit a partition to not having more than nn processor capacity at the expense of other active partitions.

► The shared processor pool. There can be one or up to 64 shared processor pools defined. Multiple partitions can be assigned to use a specific shared processor pool to help manage processor resources among like partitions. Pick a shared pool list with the names of each configured shared processor pool. This list also displays the pool ID of each configured shared processor pool in parentheses. If the name of the desired shared processor pool is not available here, you must first configure the desired shared processor pool using the Shared Processor Pool Management window. Shared processor partitions use the default shared processor pool called DefaultPool by default.

► Select whether the partition will be able to access extra processing power to fill up its virtual processors above its capacity entitlement (desired processor capacity). You can specify your partition as uncapped or implicitly capped. If there is spare processing power available in the shared processor pool or other partitions are not using their entitlement, an uncapped partition that has been consuming 100% of its desired processor capacity for several seconds can have the system automatically assign additional processor units to it.

► The weight (preference or priority) relative to other partitions to receive or get back previously given up processor power, in the case of an uncapped partition. With uncapped partitions, the Hypervisor can automatically move available processor capacity among partitions.

► You can also specify the minimum, desired, and maximum number of virtual processors. Virtual processor units (whole numbers) are what the operating system normally sees and uses to dispatch tasks (threads) to.

The POWER Hypervisor calculates a partition's processing power based on minimum, desired, and maximum values, processing mode, and also based on other active partitions' requirements. The actual entitlement is never smaller than the processing unit's desired value but can exceed that value in the case of an uncapped partition and up to the number of virtual processors allocated.

A partition can be defined with a processor capacity as small as 0.10 processing units. This represents 1/10th of a physical processor. Each physical processor can be shared by up to 10 shared processor partitions and the partition's entitlement can be incremented fractionally by as little as 1/100th 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.

This system supports up to a 4-core configuration (8203-E4A); 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 customer-specific application workload demands. For example, it would not be prudent to assign 10 processing units to a partition than runs CPU-intensive applications.

IBM provides the performance sizing tool - the IBM Systems Workload Estimator (WLE) to assist you in sizing your partitions correctly. WLE can be found at:

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

Additional information about virtual processors:

► A virtual processor can be either *running* (dispatched) on a physical processor, or *standby,* waiting for a physical processor to became available.

► Virtual processors do not introduce any additional abstraction level; they really are only a dispatch entity. When running on a physical processor, virtual processors run at the same speed as the physical processor.

► Each partition's profile defines CPU entitlement that determines how much processing power any given partition should receive. The total sum of CPU entitlement of all partitions cannot exceed the number of available physical processors in a shared processor pool.

► The number of virtual processors can be changed dynamically through a dynamic LPAR operation.

3.2.3 Processing mode

When you create a logical partition you can assign entire processors for dedicated use, or you can assign partial processor units from a shared processor pool. You can also specify that a partition set up to use dedicated processors can give up processor capacity when the partition's processor capacity is not being used.

This setting will define the processing mode of the logical partition. Figure 3-1 shows a diagram to illustrate the concepts discussed in this section.
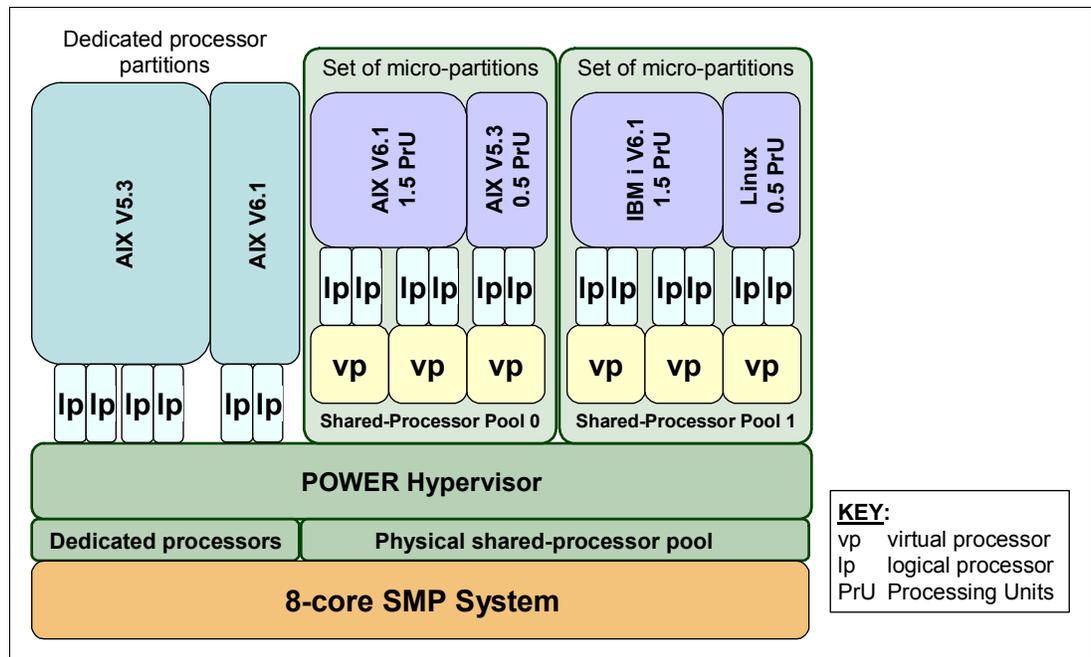

Figure 3-1   Logical partitioning concepts

The simultaneous multithreading feature in the POWER6 processor core allows the core to execute instructions from two independent software threads simultaneously. To support this feature we use the concept of *logical processors*. The operating system (AIX, IBM i, or Linux)

sees one physical processor as two logical processors if the simultaneous multithreading feature is on.

It can be turned off and on dynamically while the operating system is executing, though we recommend for almost all application environments to leave the default of "on" unchanged. For AIX, use the `smtctl` command. For IBM i use the Change System Value (CHGSYSVAL) command for system value QPRCMLTTSK. IBM i offers 3 settings: off, on, and system-controlled. The IBM i default is "system-controlled," which enables the system to normally run in "on" mode, but be able to temporarily set to "off" mode if system monitoring of task dispatching indicates performance may be improved by the *off* status, and then later change back to *on* status.

If simultaneous multithreading is off, then each physical processor is presented as one logical processor and thus only one thread at a time is executed on the physical processor.

Dedicated mode

In dedicated mode, physical processors are assigned as a whole to a partition. This is most advantageous when work in the partition is performance critical and must have that processor capacity any time work is performed.

If that partition spends a lot of time not doing any work, you could have situations where processor capacity is idle and not available for used by other partitions that could take advantage of the unused processor capacity.

You should consider specifying other processor sharing modes, as described in the following topics. Know, however, that even if you give "this partition now sharing its processor capacity" a high weight value (high priority) it will take the system a second to detect the need for processor capacity, so processing of a task that now has work to do could be delayed before sufficient capacity is "returned" to the partition.

Shared dedicated mode

On POWER6 servers, you can configure dedicated partitions to become processor donors for idle processors they own, allowing for the donation of spare CPU cycles from dedicated processor partitions to a Shared Processor Pool. The dedicated partition maintains absolute priority for dedicated CPU cycles. Enabling this feature can help to increase system utilization without compromising the computing power for critical workloads in a dedicated processor.

It is best to pilot these settings in production mode. The modes can be changed real time while the partitions are active.

Shared mode

In shared mode, logical partitions use virtual processors to access fractions of physical processors. Shared partitions can define any number of virtual processors (the maximum number is 10 times the number of processing units assigned to the partition). From the POWER Hypervisor point of view, virtual processors represent dispatching objects. The POWER Hypervisor dispatches virtual processors to physical processors according to a partition's processing units entitlement. One Processing Unit represents one physical processor's processing capacity. At the end of the POWER Hypervisor's dispatch cycle (10 ms), all partitions should receive total CPU time equal to their processing unit entitlement. The logical processors are defined on top of virtual processors. So, even with a virtual processor, the concept of logical processors exists and the number of logical processors depends on whether the simultaneous multithreading is turned on or off.

Originally only one shared processor pool was supported. Starting with POWER6 multiple shared processor pools can be defined. Specific partitions can be assigned to use each pool.

This enables you to provide more control over which partitions are competing with each other for the same processor capacity. Multiple shared pools might also help you manage to lower the number of operating system licenses required in a very dynamic uncapped partitions environment.

3.3 PowerVM

The PowerVM platform is the family of technologies, capabilities, and offerings that deliver industry-leading virtualization on IBM Power Systems servers. PowerVM is the new umbrella branding term for Power Systems Virtualization (Logical Partitioning, Micro-Partitioning, Hypervisor, Virtual I/O Server, Advanced Power Virtualization, Live Partition Mobility, Workload Partitions, and so forth). As with the predecessor branding term Advanced Power Virtualization, PowerVM is a combination of hardware enablement and value-added software. The following sections discuss the licensed features of the three editions of PowerVM.

3.3.1 PowerVM editions

PowerVM is available in three editions: Express, Standard, and Enterprise Edition. Upgrades from one PowerVM edition to another one are possible and are completely non disruptive. The upgrade does not even require the installation of additional software; the client just enters a key code in the hypervisor to unlock the next level of function.

Table 3-2 outlines the functional elements of each edition of PowerVM.

Table 3-2   PowerVM capabilities

PowerVM capability	PowerVM Express Edition (#7983)	PowerVM Standard Edition (#8506)	PowerVM Enterprise Edition (#8507)
Micro-partitions	Yes	Yes	Yes
Virtual I/O Server	Yes	Yes	Yes
Integrated Virtualization Manager	Yes	Yes	Yes
Shared Dedicated Capacity	Yes	Yes	Yes
Multiple Shared-Processor Pools	No	Yes	Yes
PowerVM Lx86	Yes	Yes	Yes
PowerVM Live Partition Mobility	No	No	Yes
Maximum # Logical Partitions	2 per system + 1 Virtual I/O Server	Up to 10 per core	Up to 10 per core

For details see 3.3.8, "Operating system support for PowerVM" on page 151 in this publication. Additional information is available from:

► IBM Power Systems PowerVM Editions Web site:

  http://www-03.ibm.com/systems/power/software/virtualization/editions/index.html

► *PowerVM Virtualization on IBM System p Managing and Monitoring*, SG24-7590, at:

  http://www.ibm.com/redbooks

► IBM Systems Hardware Information Center at:

  http://publib.boulder.ibm.com/infocenter/systems/scope/hw/index.jsp

Search on "PowerVM."

> **Note:** Systems ordered with the PowerVM Express Edition can only be managed through the Integrated Virtualization Manager. On the other hand, systems ordered with the PowerVM Standard or Enterprise Edition can be managed through the Hardware Management Console or the Integrated Virtualization Manager.

3.3.2 Virtual I/O Server

The Virtual I/O Server is part of all PowerVM Editions. It is a special purpose partition that allows the sharing of physical resources between logical partitions to allow more efficient utilization (for example, consolidation). In this case the Virtual I/O Server owns the physical resources (SCSI, Fibre Channel, network adapters, and optical devices) and allows client partitions to share access to them, thus minimizing the number of physical adapters in the system. The Virtual I/O Server eliminates the requirement that every partition owns a dedicated network adapter, disk adapter, and disk drive. The Virtual I/O Server supports OpenSSH for secure remote logins. It also provides a firewall for limiting access by ports, network services, and IP addresses. Figure 3-2 shows an overview of a Virtual I/O Server configuration.
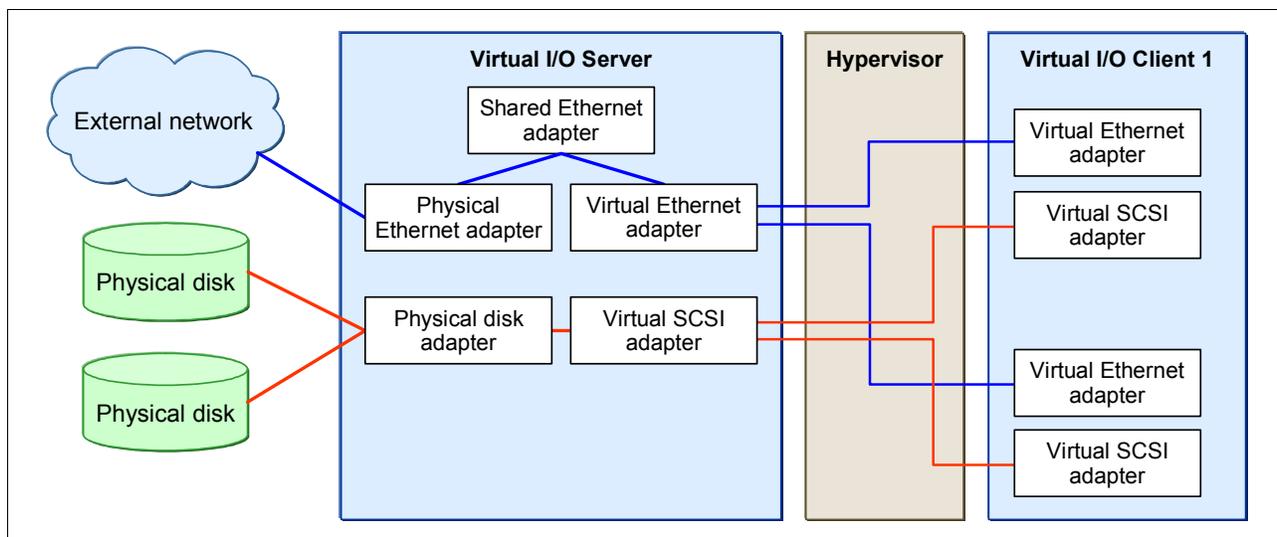
Figure 3-2   Architectural view of the Virtual I/O Server

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 any PowerVM edition. 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. 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) can be used to connect a physical Ethernet network to a virtual Ethernet network. The Shared Ethernet Adapter provides this access by connecting the internal Hypervisor VLANs with the VLANs on the external switches. Because the Shared

Ethernet Adapter processes packets at layer 2, the original MAC address and VLAN tags of the packet are visible to other systems on the physical network. IEEE 802.1 VLAN tagging is supported.

The Shared Ethernet Adapter also provides the ability for several client partitions to share one physical adapter. Using an SEA, you can connect internal and external VLANs using a physical adapter. The Shared Ethernet Adapter service can only be hosted in the Virtual I/O Server, not in a general purpose AIX or Linux partition, and acts as a layer-2 network bridge to securely transport network traffic between virtual Ethernet networks (internal) and one or more (EtherChannel) physical network adapters (external). These virtual Ethernet network adapters are defined by the POWER Hypervisor on the Virtual I/O Server.

> **Tip:** A Linux partition can provide bridging function as well, by using the `brctl` command.

Figure 3-3 shows a configuration example of an SEA with one physical and two virtual Ethernet adapters. An SEA can include up to 16 virtual Ethernet adapters on the Virtual I/O Server that share the same physical access.
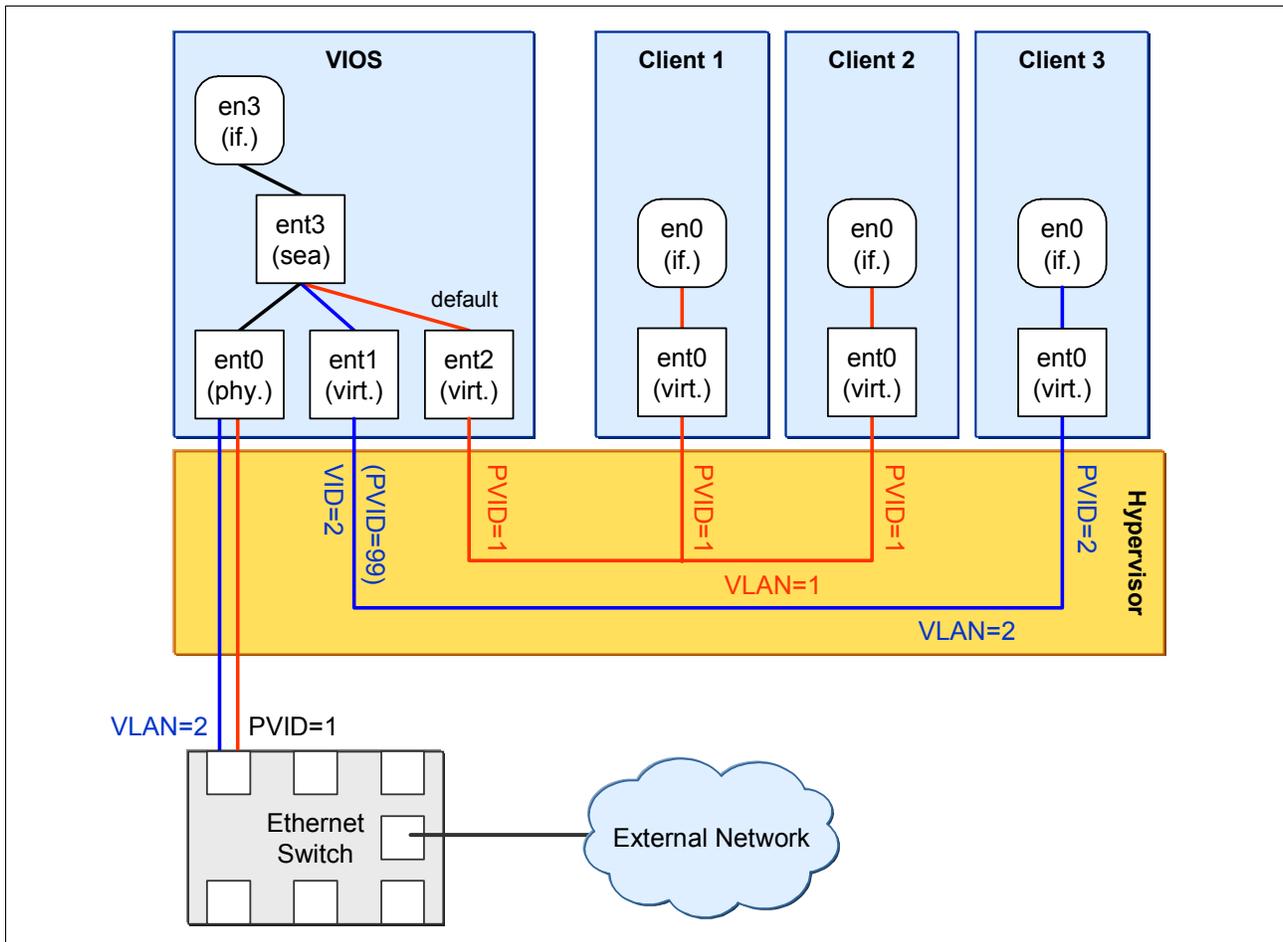

Figure 3-3   Architectural view of a Shared Ethernet Adapter

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 among 320 internal VLANs. The number of shared

Ethernet adapters that can be set up in a Virtual I/O server partition is limited only by the resource availability because there are no configuration limits.

Unicast, broadcast, and multicast are supported, so protocols that rely on broadcast or multicast, such as Address Resolution Protocol (ARP), Dynamic Host Configuration Protocol (DHCP), Boot Protocol (BOOTP), and Neighbor Discovery Protocol (NDP) can work across an SEA.

> **Note:** A Shared Ethernet Adapter does not need to have an IP address configured to be able to perform the Ethernet bridging functionality. It is very convenient to configure IP on the Virtual I/O Server. This is because the Virtual I/O Server can then be reached by TCP/IP, for example, to perform dynamic LPAR operations or to enable remote login. This can be done either by configuring an IP address directly on the SEA device, or on an additional virtual Ethernet adapter in the Virtual I/O Server. This leaves the SEA without the IP address, allowing for maintenance on the SEA without losing IP connectivity in case SEA failover is configured.

For a more detailed discussion about virtual networking, see:

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

Virtual SCSI

Virtual SCSI is used to refer to a virtualized implementation of the SCSI protocol. Virtual SCSI is based on a client/server relationship. The Virtual I/O Server logical partition owns the physical resources and acts as the server or, in SCSI terms, the target device. The client logical partitions access the virtual SCSI backing storage devices provided by the Virtual I/O Server as clients.

The virtual I/O adapters (virtual SCSI server adapter and a virtual SCSI client adapter) are configured using an HMC or through the 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. The provisioning of virtual disk resources is provided by the Virtual I/O Server.

Physical disks presented to the Virtual I/O Server can be exported and assigned to a client partition in a number of different ways:

► The entire disk is presented to the client partition.

► The disk is divided into several logical volumes, which can be presented to a single client or multiple different clients.

► As of Virtual I/O Server 1.5, files can be created on these disks and file backed storage devices can be created.

The Logical volumes or files can be assigned to different partitions. Therefore, virtual SCSI enables sharing of adapters as well as disk devices.

Figure 3-4 on page 144 shows an example where one physical disk is divided into two logical volumes by the Virtual I/O Server. Each of the two client partitions is assigned one logical volume, which is then accessed through a virtual I/O adapter (VSCSI Client Adapter). Inside the partition, the disk is seen as a normal hdisk.
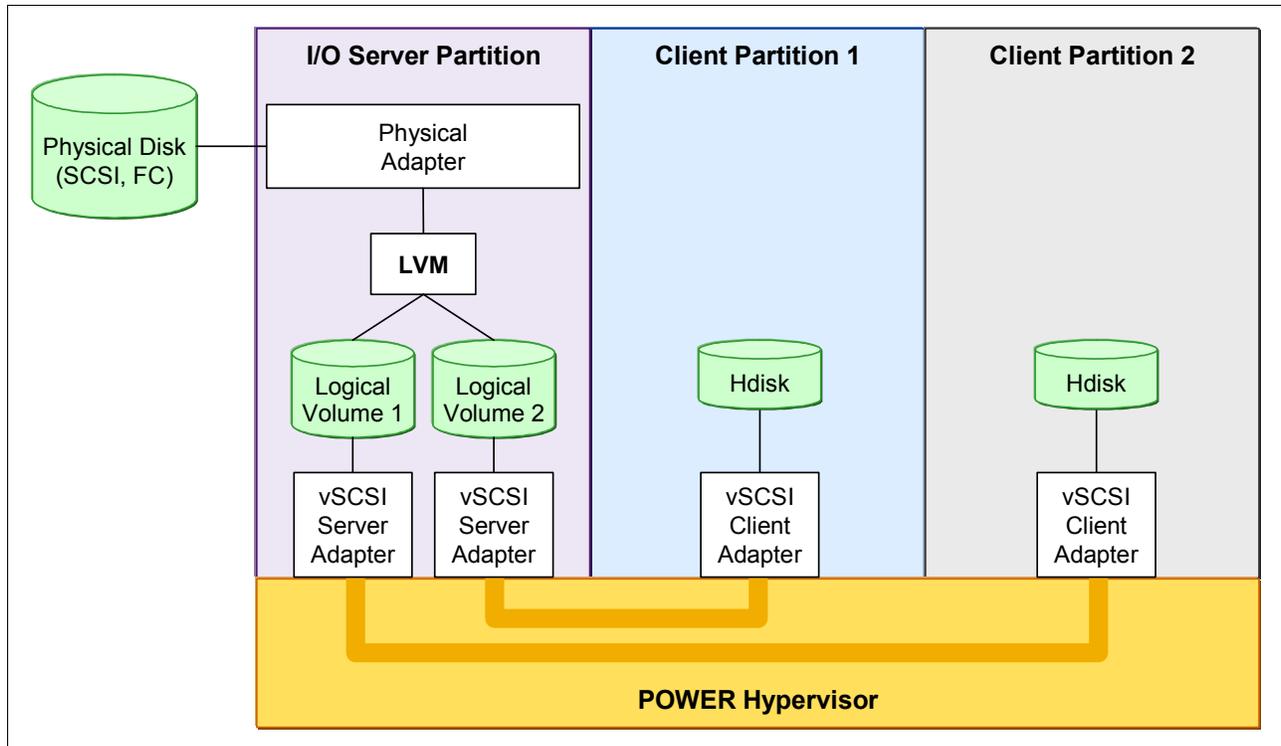
Figure 3-4   Architectural view of virtual SCSI

At the time of writing, virtual SCSI supports Fibre Channel, parallel SCSI, iSCSI, SAS, SCSI RAID devices and optical devices, including DVD-RAM and DVD-ROM. Other protocols such as SSA and tape devices are not supported.

For more information about the specific storage devices supported for Virtual I/O Server, see:

http://www14.software.ibm.com/webapp/set2/sas/f/vios/documentation/datasheet.html

Virtual I/O Server functions

Virtual I/O Server includes a number of features, including the following monitoring solutions:

▶ Support for Live Partition Mobility on POWER6 processor-based systems with the PowerVM Enterprise Edition. More information about Live Partition Mobility is in 3.3.5, "PowerVM Live Partition Mobility" on page 147.

▶ Support for virtual SCSI devices backed by a file. These are then accessed as standard SCSI-compliant LUNs.

▶ Virtual I/O Server Expansion Pack with additional security functions like Kerberos (Network Authentication Service for users and Client and Server Applications), SNMP v3 (Simple Network Management Protocol) and LDAP (Lightweight Directory Access Protocol client functionality).

▶ System Planning Tool (SPT) and Workload Estimator are designed to ease the deployment of a virtualized infrastructure. More on the System Planning Tool in section 3.5, "System Planning Tool" on page 154.

▶ IBM Systems Director and a number of preinstalled Tivoli agents are included, like Tivoli Identity Manager (TIM), to enable easy integration into an existing Tivoli Systems Management infrastructure, and Tivoli Application Dependency Discovery Manager (ADDM), which creates and maintains automatically application infrastructure maps including dependencies, change histories, and deep configuration values.

- ► vSCSI eRAS
- ► Additional Command Line Interface (CLI) statistics in `svmon`, `vmstat`, `fcstat`, and `topas`.
- ► Monitoring solutions to help manage and monitor the Virtual I/O Server and shared resources. New commands and views provide additional metrics for memory, paging, processes, Fibre Channel HBA statistics, and virtualization.
- ► The Integrated Virtualization Manager now supports dedicated I/O, as discussed in the next section.

For more information on the Virtual I/O Server and its implementation, refer to *PowerVM Virtualization on IBM System p: Introduction and Configuration, Fourth Edition*, SG24-7940.

3.3.3 Integrated Virtualization Manager

In order to ease virtualization technology adoption in any IBM System p environment, IBM has developed Integrated Virtualization Manager (IVM), a simplified hardware management solution for a single system that inherits some of the Hardware Management Console features, thus avoiding the necessity of a dedicated control workstation. The Integrated Virtualization Manager runs on top of the Virtual I/O Server.

Prior to VIOS/IVM V1.5 all the physical resources were allocated to the Virtual I/O Server logical partition; however, with Version 1.5 there is also support for dedicated physical I/O resources to logical partitions other than the Virtual I/O Server partition.

The IVM solution enables the administrator to reduce system setup time. IVM is targeted at small and medium systems. Hardware management is done remotely through a Web browser based interface.

IVM does not require network connectivity with the system's service processor. A specific device named the *Virtual Management Channel (VMC)* has been developed on the Virtual I/O Server to enable a direct POWER Hypervisor configuration without requiring additional network connections to be set up. This device is activated by default when Virtual I/O Server is installed as the first partition on a system without an HMC console.

IVM can be used to complete the following tasks:

- ► Create and dynamically manage logical partitions (dynamic LPAR reconfiguration)
- ► 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

Among the major differences between IVM and an HMC-managed system are that on IVM:

- ► There is only one profile per partition.
- ► Each LPAR can have a maximum of one Virtual SCSI adapter assigned.
- ► IVM supports a single Virtual I/O Server to support all your mission-critical production needs.

Despite those considerations, IVM provides advanced virtualization functionality without the need for an extra-cost workstation. For more information about IVM functionality and best practices, see *Integrated Virtualization Manager on IBM System p5*, REDP-4061.

> **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 you should never connect an HMC to a working IVM system. If you want to migrate an environment from IVM to HMC, you have to rebuild the configuration setup manually.

3.3.4 PowerVM Lx86

The IBM PowerVM Lx86 feature creates a virtual x86 Linux application environment on POWER processor-based systems, so most 32-bit x86 Linux applications can run without requiring clients or ISVs to recompile the code. This brings new benefits to organizations that want the reliability and flexibility of consolidating (through virtualization) on Power Systems and use applications that have not yet been ported to the platform.

PowerVM Lx86 dynamically translates x86 instructions to Power Architecture instructions, operating much like the Just-in-time compiler (JIT) in a Java system. The technology creates an environment in which the applications being translated run on the new target platform, in this case Linux. This environment encapsulates the application and runtime libraries and runs them on the Linux operating system kernel. These applications can be run side by side with POWER native applications on a single system image and do not require a separate partition.

The Linux x86 application environment is illustrated in Figure 3-5.
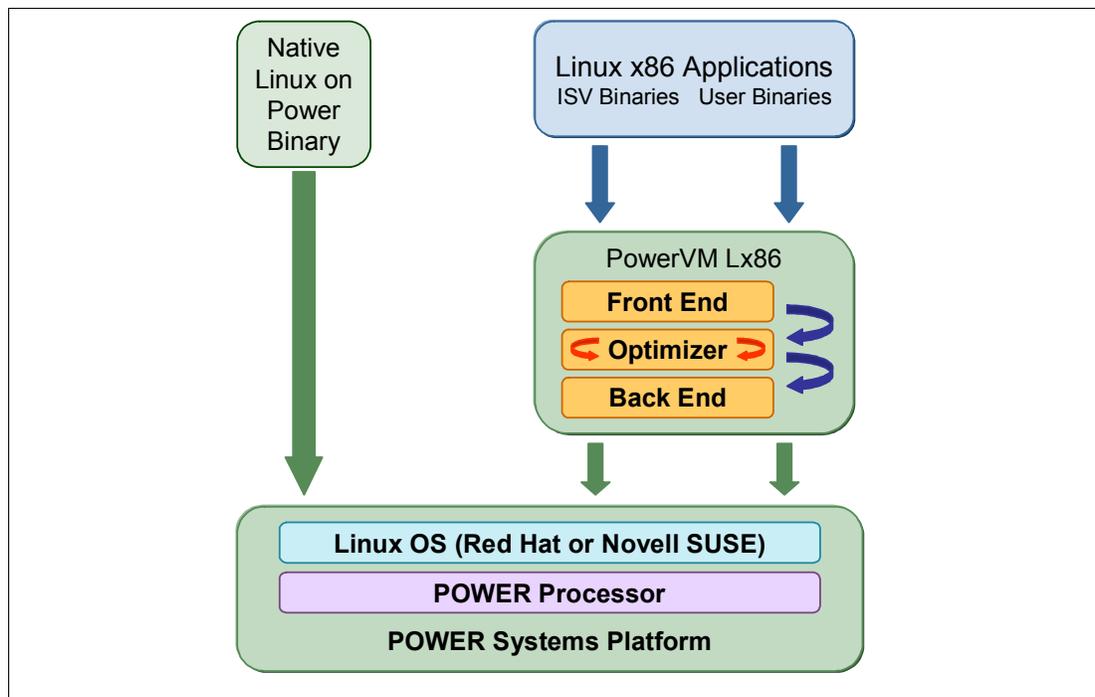

Figure 3-5   Diagram of the Linux x86 Application Environment

Supported Operating Systems

PowerVM Lx86 version 1.1 will support the following Linux operating systems:

- ► Red Hat Enterprise Linux 4 (RHEL 4) for POWER version 4.4 and 4.5. Also, x86 Linux applications running on RHEL 4.3 are supported.
- ► SUSE Linux Enterprise Server 9 (SLES 9) for POWER Service Pack 3.
- ► SUSE Linux Enterprise Server 10 (SLES 10) for POWER Service Pack 1.

Note:

- ► PowerVM Lx86 is supported under the VIOS Software Maintenance Agreement (SWMA).
- ► When using PowerVM Lx86 on an IBM System p POWER6 processor-based system only SLES 10 with SP1 and RHEL 4.5 are supported.
- ► Make sure the x86 version is the same as your Linux version. Do not try to use any other version because it is unlikely to work. One exception is Red Hat Enterprise Linux, with which both the Advanced Server and Enterprise Server option at the correct release will work.

As stated previously, PowerVM Lx86 runs most x86 Linux applications, but PowerVM Lx86 cannot run applications that:

- ► Directly access hardware devices (for example, graphics adapters)
- ► Require nonstandard kernel module access or use kernel modules not provided by the Linux for Power Systems operating system distribution
- ► Do not use only the Intel® IA-32 instruction set architecture as defined by the 1997 Intel Architecture Software Developer's Manual consisting of Basic Architecture (Order Number 243190), Instruction Set Reference Manual (Order Number 243191) and the System Programming Guide (Order Number 243192) dated 1997
- ► Do not run correctly on Red Hat Enterprise Linux 4 starting with version 4.3 or Novell SUSE Linux Enterprise Server (SLES) 9 starting with version SP3 or Novell SLES 10
- ► Are x86 Linux-specific system administration or configuration tools

For more information about PowerVM Lx86, refer to *Getting Started with PowerVM Lx86*, REDP-4298.

3.3.5 PowerVM Live Partition Mobility

PowerVM Live Partition Mobility allows you to move a running logical partition, including its operating system and running applications, from one system to another without any shutdown and without disrupting the operation of that logical partition. Inactive partition mobility allows you to move a powered off logical partition from one system to another.

Partition mobility provides systems management flexibility and improves system availability by enabling you to:

- ► Avoid planned outages for hardware or firmware maintenance by moving logical partitions to another server and then performing the maintenance. Live partition mobility can help lead to zero downtime maintenance because you can use it to work around scheduled maintenance activities.

- ► Avoid downtime for a server upgrade by moving logical partitions to another server and then performing the upgrade. This allows your end users to continue their work without disruption.
- ► Preventive failure management: If a server indicates a potential failure, you can move its logical partitions to another server before the failure occurs. Partition mobility can help avoid unplanned downtime.
- ► Achieve server optimization by:
  - – Consolidating workloads running on several small, under-used servers onto a single large server.
  - – Moving workloads from server to server to optimize resource use and workload performance within your computing environment. With active partition mobility, you can manage workloads with minimal downtime.

Mobile partition operating system requirements

The operating system running in the mobile partition must be AIX or Linux. The Virtual I/O Server logical partition must be at least at the 1.5 release level. However, the Virtual I/O Server partition itself cannot be migrated. The operating system must be at one of the following levels:

- – AIX 5L V5.3 with 5300-07 Technology Level or later
- – AIX V6.1 or later
- – Red Hat Enterprise Linux Version V5.1 or later
- – SUSE Linux Enterprise Services 10 (SLES 10) Service Pack 1 or later

Previous versions of AIX and Linux can participate in inactive partition mobility, if the operating systems support virtual devices and IBM System p POWER6 processor-based systems.

Source and destination system requirements

The source partition must be one that only has virtual devices. If there are any physical devices in its allocation, they must be removed before the validation or migration is initiated.

The hypervisor must support the Partition Mobility functionality, also called the migration process. POWER 6 processor-based hypervisors have this capability and firmware must be at firmware level eFW3.2 or later. Source and destination systems could have different firmware levels, but they must be compatible with each other.

The Virtual I/O Server on the source system provides the access to the client's resources and must be identified as a Mover Service Partition (MSP). The VASI device, Virtual Asynchronous Services Interface (VASI) allows the mover service partition to communicate with the hypervisor; it is created and managed automatically by the HMC and will be configured on both the source and destination Virtual I/O Servers designated as the mover service partitions for the mobile partition to participate in active mobility. Other requirements include a similar Time of Day on each server, systems should not be running on battery power, shared storage (external hdisk with reserve_policy=no_reserve), and all logical partitions should be on the same open network with RMC established to the HMC.

The HMC is used to configure, validate, and orchestrate. You use the HMC to configure the Virtual I/O Server as an MSP and to configure the VASI device. An HMC wizard validates your configuration and identifies things that will cause the migration to fail. During the migration, the HMC controls all phases of the process.

For more information about Live Partition Mobility and how to implement it, refer to *IBM System p Live Partition Mobility*, SG-24-7460.

3.3.6 PowerVM AIX 6 workload partitions

Workload partitions provide a way for clients to run multiple applications inside the same instance of an AIX operating system while providing security and administrative isolation between applications. Workload partitions complement logical partitions and can be used in conjunction with logical partitions and other virtualization mechanisms, if desired. Workload partitions (WPAR) is a software-based virtualization capability of AIX V6 that can improve administrative efficiency by reducing the number of AIX operating system instances that must be maintained, can increase the overall utilization of systems by consolidating multiple workloads on a single system, and is designed to improve cost of ownership.

The use of workload partitions is optional; therefore, programs will run as before if run in the Global environment (AIX instance). This Global environment owns all the physical resources (like adapters, memory, disks, processors) of the logical partition.

> **Note:** Workload partitions are only supported with AIX V6.

Workload partitions are separate regions of application space, and therefore they allow users to create multiple software-based partitions on top of a single AIX instance. This approach enables high levels of flexibility and capacity utilization for applications executing heterogeneous workloads, and simplifies patching and other operating system maintenance tasks.

There are two types of workload partitions:

► System workload partitions: These are autonomous virtual system environments with their own private root file systems, users and groups, login, network space, and administrative domain. This type of partition within the operating system isolates runtime resources such as memory, CPU, user information, or file system to specific application processes. Each System workload partition has its own unique set of users, groups, and network addresses. It is integrated with Role-Based Access control (RBAC). Inter-process communication for a process in a workload partition is restricted to those processes in the same workload partition. The systems administrator accesses the workload partition via the administrator console or via regular network tools such as telnet or ssh. The system workload partition is removed only when requested.

► Application workload partitions: These are lightweight workload partitions because there are no system services involved, there is no file system isolation since they use the global environment system file system, and Telnet is not supported but access through console login is available. Once the application process or processes are finished the workload partition is stopped.

For a detailed discussion about workload partitions concepts and functions, refer to *Introduction to Workload Partition Management in IBM AIX Version 6.1*, SG24-7431.

3.3.7 PowerVM AIX 6 Workload Partition Manager

IBM PowerVM AIX 6 Workload Partition Manager (WPAR Manager) is a platform management solution that provides a centralized point of control for managing workload partitions (WPARs) across a collection of managed systems running AIX. It is an optional product designed to facilitate the management of WPARs as well as to provide advanced

features such as policy-based application mobility for automation of workload partitions relocation based on current performance state.

The workload partition manager is an intuitive GUI-based tool designed to provide a centralized interface for administration of WPAR instances across multiple systems. By deploying the workload partitions manager, users are able to take full advantage of workload partitions technology by leveraging the following features:

► Basic life cycle management

 Create, start, stop, and delete WPAR instances

► Manual WPAR mobility

 User-initiated relocation of WPAR instances

► Creation and administration of mobility policies

 User-defined policies governing automated relocation of WPAR instances based on performance state

► Creation of compatibility criteria on a per WPAR basis

 User-defined criteria based on compatibility test results gathered by the WPAR Manager

► Administration of migration domains

 Creation and management of server groups associated to specific WPAR instances that establish which servers would be appropriate as relocation targets

► Server profile ranking

 User-defined rankings of servers for WPAR relocation based on performance state

► Reports based on historical performance

 Performance metrics gathered by WPAR Manager for both servers and WPAR instances

► Event logs and error reporting

 Detailed information related to actions taken during WPAR relocation events and other system operations

► Inventory and automated discovery

 Complete inventory of WPAR instances deployed on all servers with WPAR Manager Agents installed, whether created by the WPAR Manager or through the command line interface (CLI) on the local system console

Note: The IBM PowerVM Workload Partition Manager for AIX Version 6 is a separate, optional orderable feature (#5765-WPM).

3.3.8 Operating system support for PowerVM

Table 3-3 lists AIX, IBM i, and Linux for Power Systems operating system support for PowerVM capabilities. This is based on:

http://www–03.ibm.com/systems/power/software/virtualization/editions/index.html

Table 3-3   PowerVM features supported by AIX, IBM i, and Linux

Feature	AIX V5.3	AIX V6.1	IBM i 5.4 with LIC 5.4.5	IBM i 6.1	RHEL V4.5 for Power	RHEL V5.1 for Power	SLES V10 SP1 for Power
Dynamic LPAR operations[1]	Y	Y	Y	Y	Y	Y	Y
Capacity Upgrade on Demand[2]	Y	Y	Y	Y	Y	Y	Y
Micro-Partitioning	Y	Y	Y	Y	Y	Y	Y
Shared Dedicated Processor Capacity	Y	Y	N	Y	N	Y	Y
Multiple Shared Processor Pools	Y	Y	N	Y	N	Y	Y
Virtual I/O Server	Y	Y	N	Y[5]	Y	Y	Y
IVM	Y	Y	Y	Y	Y	Y	Y
Virtual SCSI	Y	Y	Y[3]	Y[6]	Y	Y	Y
Virtual Ethernet	Y	Y	Y[3]	Y[6]	Y	Y	Y
Live Partition Mobility	Y	Y	N	Y	N	Y	Y
Workload Partitions	N	Y	Y[4]	Y[4]	N	N	N

1. Dynamic memory removal is not supported by Linux at the time of writing.
2. Available on selected models.
3. As a server.
4. IBM I standard support (PowerVM edition not required) includes multiple workloads and subsystems per partition. These capabilities are logically similar to those of AIX V6.1 Workload Partitions.
5. As a client.
6. As a server or client.

Table 3-4 on page 152 summarizes the PowerVM capabilities per current IBM POWER6-based processor servers and POWER5/5+ System i and System p servers. This support assumes the appropriate operating system release level has been installed. This table is meant to be used in conjunction with Table 3-3.

Additional details are on the PowerVM Web site at:

http://www.ibm.com/systems/power/software/virtualization/editions/index.html

Table 3-4 PowerVM features support on POWER6 and POWER5/5+ system models and machine types

PowerVM features by POWER6-based systems	JS12 7998-60X	JS22 7998-61X	520 8203-E4A	520 9407-M15	520 9408-M25	550 8204-EA8	550 9409-M50	570 9117-MMA	570 9406-MMA	575 9125-F2A	595 9119-FHA	System p POWER5/5+	System i POWER5/5+ [3]
Express Edition	N	N	Y	N	N	Y	N	N	N	N	N	N	N
Standard Edition	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Enterprise Edition	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	N	N
Micro-partitioning	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
VIOS	Y	Y	Y	N	Y	Y	Y	Y	Y	Y			Y[1]
IVM	Y	Y	Y	N	N	Y	N	N	N	N	N	Y[2]	N
HMC	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Lx86	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	N	N
Live Partition Mobility	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	N	N
Multiple Shared Processor Pools	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N
Shared Dedicated Processor Capacity	N	N	Y	N	Y	Y	Y	Y	Y	Y	Y	N	N

1. Dynamic memory removal is not supported by Linux at the time of writing.
2. IVM is supported on System p POWER5/POWER5+ servers and below. It is also supported on the JS21.
3. PowerVM is not required for micro-partitioning or for IBM i (i5/OS) virtual I/O support (not VIOS) to a hosted client AIX or Linux on a POWER5/5+ partition, except for the POWER5+ 9407-Model 515. PowerVM can be used on any System i POWER5/5+ model to support VIOS capabilities to a client AIX or Linux partition. IVM is not supported on any System i POWER5/5+ system.
   IBM i 6.1 or later on a POWER6 system is required for IBM i to be a client partition to host VIOS partition or a host IBM i 6.1 partition

Making effective use of the PowerVM capabilities in moderate to complex operating environments requires proper planning for use of the system's hardware resources. For additional information in this area we recommend that you review some of the many PowerVM-related articles available at the PowerVM Web site:

http://www-03.ibm.com/systems/power/software/virtualization/

Documents such as *System i and System p: Advanced POWER Virtualization Operations Guide*, SA76-0100 are especially useful. This publication contains detailed information, including sizing for performance the capabilities of functions such as the Virtual I/O Server (VIOS) partitions and communications traffic handled by the Integrated Virtual Ethernet adapter (IVE). The IVE is also known as the Host Ethernet Adapter (HEA).

The PDF of this publication is available from the IBM Systems Hardware Information Center at:

http://publib.boulder.ibm.com/infocenter/systems/scope/hw

- ► Search with Advanced POWER Virtualization Operations Guide, then select PowerVM Editions → PowerVM Editions Operations Guide and click the link to the PDF.

We also suggest reviewing IBM Redbooks publication *IBM PowerVM Virtualization Managing and Monitoring*, SG24-7590.

3.4 Active Memory Sharing

PowerVM is enhanced with PowerVM Active Memory™ Sharing, an advanced memory virtualization technology that intelligently flows memory from one partition to another for increased utilization and flexibility of memory usage. With this memory virtualization enhancement for IBM POWER6 processor-based servers, IBM i, AIX, and Linux partitions can share a pool of memory and have PowerVM automatically allocate the memory based on the partition's workload demands. Active Memory Sharing is provided with PowerVM Enterprise Edition only.

PowerVM Active Memory Sharing supports over-commitment of logical memory with overflow going to a PowerVM Virtual I/O Server (VIOS)-managed paging device. For redundancy, two VIOS partitions can be configured to the shared memory pool to provide access to the paging devices. In addition, AIX and Linux partitions that use Active Memory Sharing can leverage Live Partition Mobility to help improve resiliency and avoid planned downtime.

Active Memory Sharing minimum requirements:

- ► POWER6 processor-based system
- ► eFW 3.4.2
- ► Hardware Management Console V7R3.4.0 Service Pack 2
- ► Integrated Virtualization Manager 2.1.1
- ► Virtual I/O Server 2.1.1
- ► IBM i 6.1 or AIX 6.1 with 6100-03 Technology Level or SUSE Linux Enterprise Server 11

> **Note:** As of 15th May 2009, IBM PowerVM has been enhanced to provide additional flexibility and manageability. Additional PowerVM enhancements include:
>
> - ► Active Memory Sharing
> - ► PowerVM VIOS management enhancements
> - ► N-port ID Virtualization (NPIV) support for SUSE Linux Enterprise Server 11 and the IBM Power 595
> - ► Virtual tape support for IBM i and SUSE Linux Enterprise Server 11
> - ► VIOS support for solid state drives (SSD), DS5000, XIV storage systems
> - ► PowerVM Lx86 enhancements

Figure 3-6 on page 154 shows several examples of maximizing memory usage efficiency in Power 520 and Power 550 using PowerVM Active Memory Sharing feature.
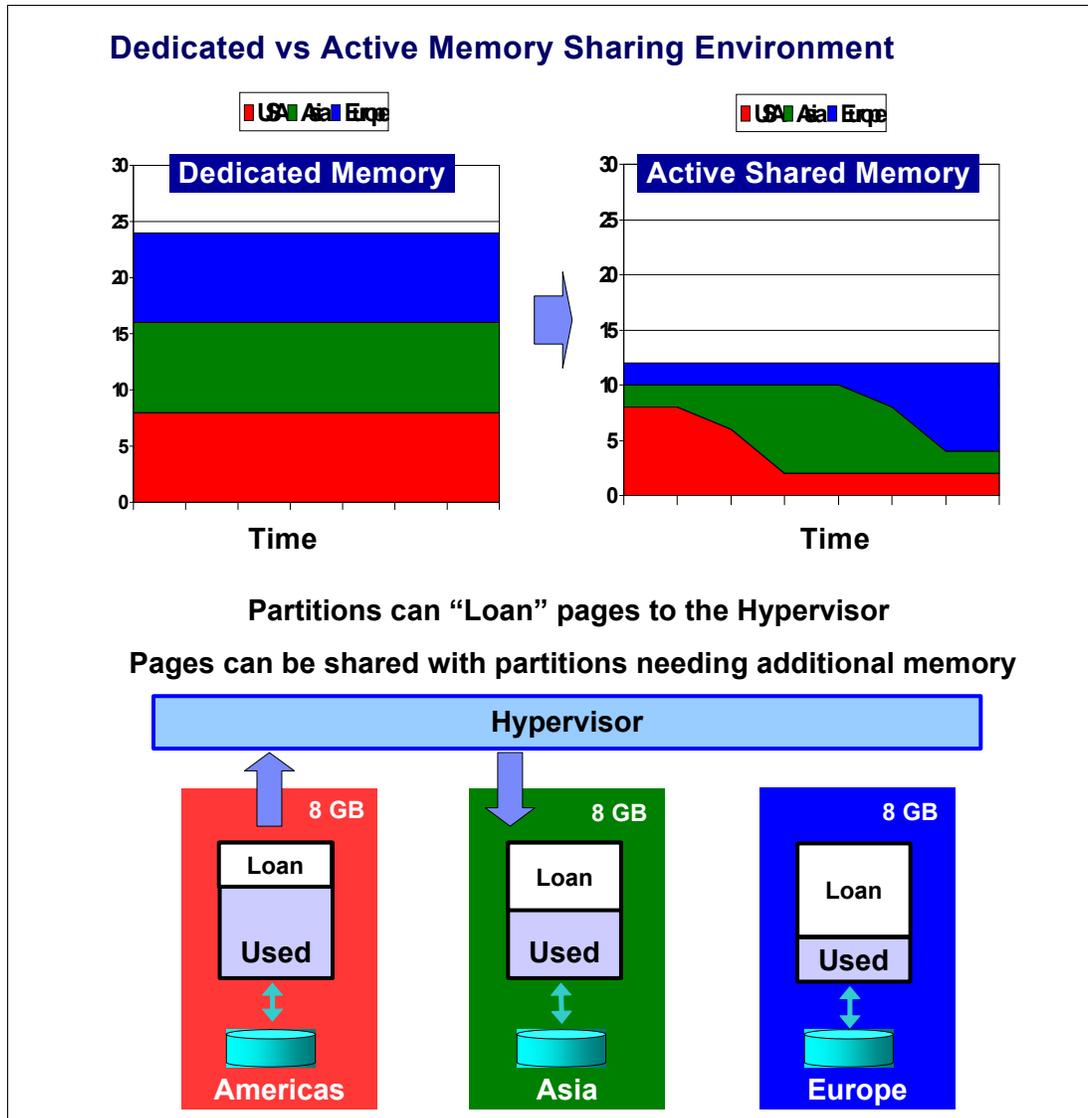
Figure 3-6   Typical examples of PowerVM Memory Active Sharing feature

3.5 System Planning Tool

The IBM System Planning Tool (SPT) helps you select your hardware configuration:
processor speeds and activated processor cores, main memory size and memory cards, and
the necessary I/O hardware-elated elements such I/O adapter, supported I/O devices, I/O
loop adapters, and I/O enclosures. It can be used for a single partition system. Its strengths,
however, are best demonstrated when configuring a system with multiple partitions and
multiple operating systems.

The resulting output of your design is called a *system plan*, which is stored in a .sysplan file.
This file can contain plans for a single system or multiple systems. The .sysplan file can be
used:

► To create reports
► As input to the Configuration file, which can be input to the IBM sales configurator tool
  described later in this section

► To create and deploy partitions on your systems automatically

SPT is the follow on to the original IBM LPAR Validation Tool (LVT). It contains all the functions from the LVT, as well as significant functional enhancements introduced during 2007 and 2008, further integrating System i, System p, and Power Systems hardware configurations.

You can have your current configuration and your completed configuration validated to minimize additional "adjustments" when using the IBM sales configuration tool.

SPT can also be linked to the IBM Systems Workload Estimator (WLE) to help you estimate the processor capacity, memory size, and number of disk arms needed for a single system or multiple partitions on the same system.

You can create an entirely new system configuration, or you can create a system configuration based on any of the following:

► Performance data from an existing system that the new system is to replace

► Performance estimates that anticipate future workloads that you must support

► Sample systems that you can customize to fit your needs

In addition to determining your hardware configuration for partitions on a new system you can also use the SPT to determine how you can partition a system that you already have.

The SPT is a PC-based browser application designed to be run in a standalone environment on your PC workstation. The SPT can be used to plan solutions for the following IBM systems:

► IBM Power Systems

► System p5 and System i5®

► OpenPower (Linux)

► iSeries® 8xx and 270 models

SPT can create two files from your configuration work (which can include any combination of multiple AIX, IBM i, and Linux partitions).

► Configuration file: The configuration file, denoted by the *.cfr* file extension, can be imported to the IBM sales configurator tool (e-Config) as input to the ordering process. Using SPT to create the configuration files helps IBM manufacturing configure your new order so that, as much as is possible, all the hardware is physically located and ready for use by the logical partitions you have configured. Using this file helps minimize moving cards to different card slots when deploying partition configurations on the delivered system.

► System plan file: When the system plan file, denoted by the *.sysplan* file extension, is on the Hardware Management Console, it can be used to deploy the partition configurations defined in the sysplan on the installed system. This file can be exported to media and used by an HMC to deploy the partition configurations stored within the sysplan file.

We recommend that you consider using the IBM Systems Workload Estimator (WLE) tool previously mentioned to "performance size" your POWER6 partitions. This tool can be accessed via the Web at:

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

This can help you determine the processor capacity, memory, and disk configuration sufficient to meet you performance objectives for each partition (or an only partition) on a system.

Based on that output, you can use the IBM System Planning Tool to configure the hardware for each partition. As part of its output, SPT also estimates POWER Hypervisor requirements as part of the memory resources required for all partitioned and non-partitioned servers.

Though not up to date with the latest SPT capabilities, a good publication to get an overview of SPT capabilities, including its integration with WLE, is *IBM System i and System p System Planning and Deployment: Simplifying Logical Partitioning, SG24-7487*, published in August 2007.

Figure 3-7 on page 156 is an SPT screen example showing the estimated Hypervisor memory requirements based on sample partition requirements.
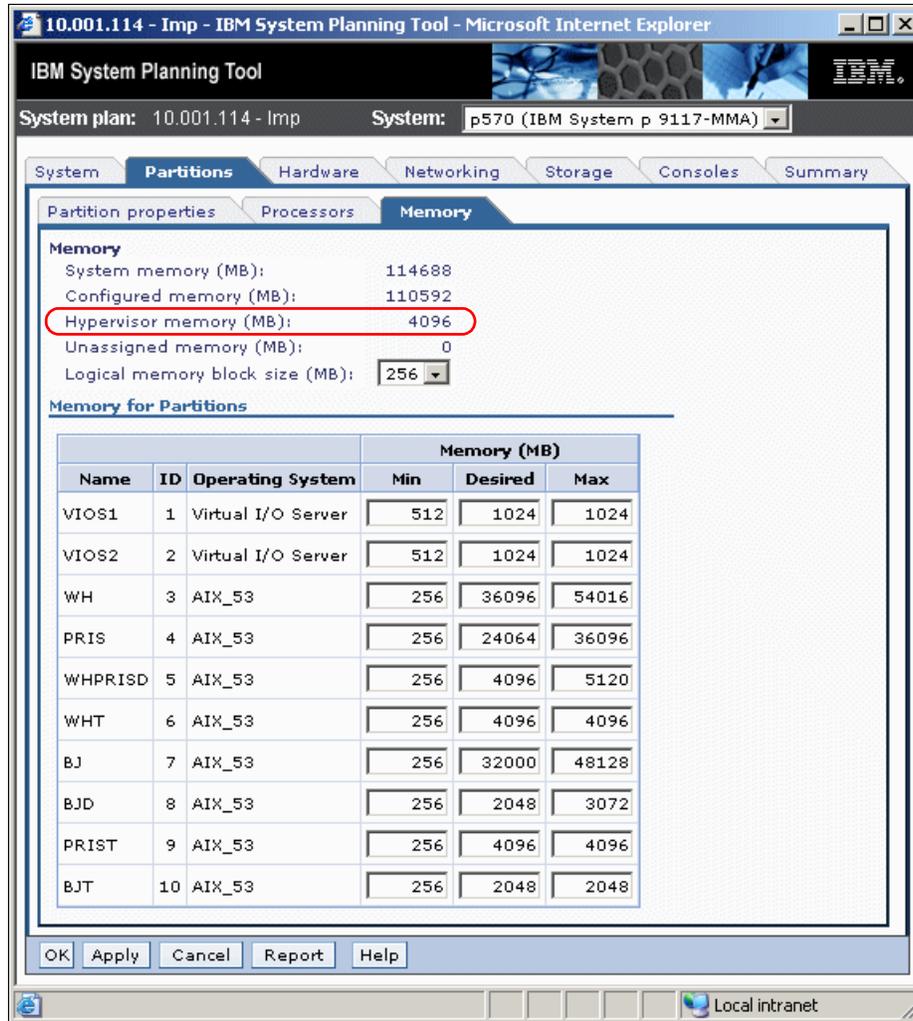
Figure 3-7   IBM System Planning Tool window showing Hypervisor memory requirements

Note: In previous releases of the SPT, you could view an HMC or IVM system plan, but you could not edit the plan. The SPT now allows you to convert an HMC or IVM system plan into a format that will allow you to edit the plan in the SPT.

SPT Version 3 (3.08.095SPT) is the latest level available at the original publication date of this publication. Note that SPT 2.0 is the last release that will support .lvt and .xml files (from the predecessor LVT tool). Users should load their old .lvt and .xml plans and save them as .sysplan files.

The SPT and its supporting documentation can be found on the IBM System Planning Tool site at:

http://www.ibm.com/systems/support/tools/systemplanningtool/

> **Note:** Consider using #0456 on new orders to specify adapter card placement within the system unit and I/O enclosures to expedite getting your LPAR configuration up and running once your system is delivered. This is where the SPT configuration file (for the order) and system plan file (for deployment) can be most useful.

4

Continuous availability and manageability

This chapter provides information about IBM Power Systems hardware design features that help lower the total cost of ownership (TCO). The advanced IBM RAS (Reliability, Availability, and Service ability) technology provides the possibility to improve your architecture's TCO by reducing unplanned down time.

IBM POWER6 processor-based systems have a number of new features that enable systems to dynamically adjust when issues arise that threaten availability. Most notably, POWER6 processor-based systems introduce the POWER6 Processor Instruction Retry suite of tools, which includes Processor Instruction Retry, Alternate Processor Recovery, Partition Availability Prioritization, and Single Processor Checkstop. Taken together, in many failure scenarios these features allow a POWER6 processor-based system to recover with no impact on a partition using a failing core.

This chapter also includes several software-based availability features that are based on the benefits available when using AIX and IBM i as the operating system. Support of these features when using Linux can vary.

4.1 Reliability

Highly reliable systems are built with highly reliable components. On IBM POWER6 processor-based systems, this basic principle is expanded upon with a clear design for reliability architecture and methodology. A concentrated, systematic, architecture-based approach is designed to improve overall system reliability with each successive generation of system offerings.

4.1.1 Designed for reliability

Systems designed with fewer components and interconnects have fewer opportunities to fail. Simple design choices such as integrating two processor cores on a single POWER chip can dramatically reduce the opportunity for system failures. In this case, a 4-core server will include half as many processor chips (and chip socket interfaces) as with a single-CPU-per-processor design. Not only does this reduce the total number of system components, it reduces the total amount of heat generated in the design, resulting in an additional reduction in required power and cooling components.

Parts selection also plays a critical role in overall system reliability. IBM uses three grades of components, with grade 3 defined as industry standard (off-the-shelf). As shown in Figure 4-1, using stringent design criteria and an extensive testing program, the IBM manufacturing team can produce grade 1 components that are expected to be 10 times more reliable than industry standard. Engineers select grade 1 parts for the most critical system components. Newly introduced organic packaging technologies, rated grade 5, achieve the same reliability as grade 1 parts.
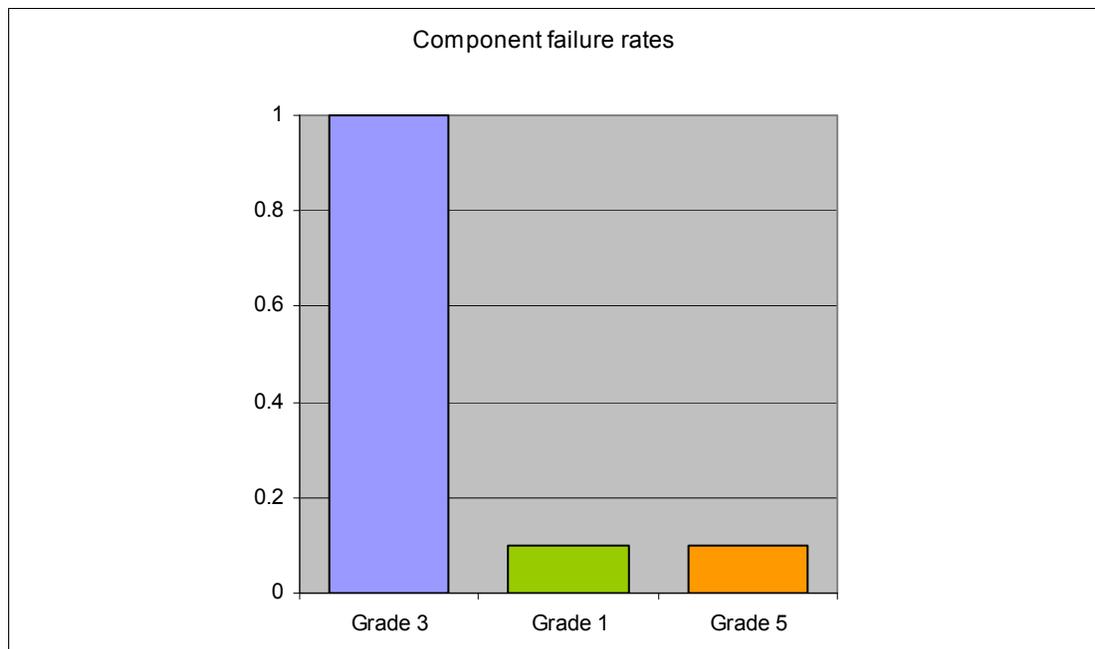

Figure 4-1   Component failure rates

4.1.2 Placement of components

Packaging is designed to deliver both high performance and high reliability. For example, the reliability of electronic components is directly related to their thermal environment, that is,

large decreases in component reliability are directly correlated with relatively small increases in temperature, POWER6 processor-based systems are carefully packaged to ensure adequate cooling. Critical system components such as the POWER6 processor chips are positioned on printed circuit cards so they receive fresh air during operation. In addition, POWER6 processor-based systems are built with redundant, variable-speed fans that can automatically increase output to compensate for increased heat in the central electronic complex.

4.1.3 Redundant components and concurrent repair

High-opportunity components, or those that most affect system availability, are protected with redundancy and the ability to be repaired concurrently.

The use of redundant parts allows the system to remain operational. Among them are:

► Redundant spare memory bits in cache, directories and main memory

► Redundant and hot-swap cooling

► Redundant and hot-swap power supplies (optional)

    AIX and OpenPower Express Product Offerings and IBM i Express Editions include two power supplies and two power cords in the configuration automatically with any new initial order. For maximum availability it is highly recommended to connect power cords from the same system to two separate Power Distribution Units (PDUs) in the rack and to connect each PDU to independent power sources. Deskside form factor power cords need to be plugged to two independent power sources in order to achieve maximum availability.

See 4.3.1, "Detecting errors" on page 171.

4.2 Hardware availability

IBM hardware and microcode capability to continuously monitor execution of hardware functions is generally described as the process of First Failure Data Capture (FFDC). This includes the strategy of Predictive Failure Analysis, which refers to the ability to track intermittent correctable errors and to vary components off-line before they reach the point of hard failure causing a system outage.

POWER6 technology builds on previous hardware capabilities in this area. In many areas this drives towards the high availability objective with hardware components that:

► Self-diagnose and self-correct during run time

► Automatically reconfigure to mitigate potential problems from suspect hardware

► Have the ability to self-heal or to automatically substitute good components for failing components

Throughout this chapter we describe IBM POWER6 technology's capabilities focused on keeping a system environment up and running. For a specific set of functions focused on detecting errors before they become serious enough to stop computing work, see 4.3.1, "Detecting errors" on page 171.

4.2.1 Partition availability priority

Before describing the array of POWER6 processor, memory, and I/O availability capabilities we first summarize partition availability priority.

Starting with POWER5 technology, partitions receive an integer "weight value" rating with the lowest priority partition rated at "0" and the highest priority partition valued at "255." The default value is set at "127" for standard partitions and "192" for Virtual I/O Server (VIOS) partitions. You can vary the priority of individual partitions.

Partition availability priorities can be set for both dedicated and shared processor partitions. The POWER Hypervisor uses the relative partition weight value among active partitions to favor higher priority partitions for processor sharing, adding and removing processor capacity, and favoring higher priority partitions for normal operation.

This same set of values are applied when providing processor-based error detection and recovery processes. For example, if a full processor availability error is encountered, the partition priority values area applied should a processor failure reduce the total system processor capacity to less than the desired capacity specified in all active partitions.

Note, the partition specifications for "required," "desired", and "maximum" capacity are also taken into account for capacity on demand options, and if total system-wide processor capacity becomes disabled due to deconfigured failed processor cores. For example, if total system-wide processor capacity is sufficient to run all partitions at least with the "required" capacity, the partitions are allowed to start or continue running. If processor capacity is insufficient to run a partition at its required value, then starting that partition has an error condition that needs to be resolved.

Messages are issued and errors are logged indicating the processor-based failure history sufficient to identify recovery actions needed to be taken.

4.2.2 General detection and deallocation of failing components

Runtime correctable or recoverable errors are monitored to determine if there is a pattern of errors. If these components reach a predefined error limit, the service processor initiates an action to deconfigure the faulty hardware, helping to avoid a potential system outage and to enhance system availability.

Persistent deallocation

To enhance system availability, a component that is identified for deallocation or deconfiguration on a POWER6 processor-based system is flagged for persistent deallocation. Component removal can occur either dynamically (while the system is running) or at boot-time (IPL), depending both on the type of fault and when the fault is detected.

In addition, runtime unrecoverable hardware faults can be deconfigured from the system after the first occurrence. The system can be rebooted immediately after failure and resume operation on the remaining stable hardware. This prevents the same faulty hardware from affecting system operation again, while the repair action is deferred to a more convenient, less critical time.

Persistent deallocation functions include:

► Processor
► Memory
► Deconfigure or bypass failing I/O adapters
► L2 cache

Note: The auto-restart (reboot) option has to be enabled from the Advanced System Manager Interface or from the Control (Operator) Panel. Figure 4-2 shows this option using the ASMI.
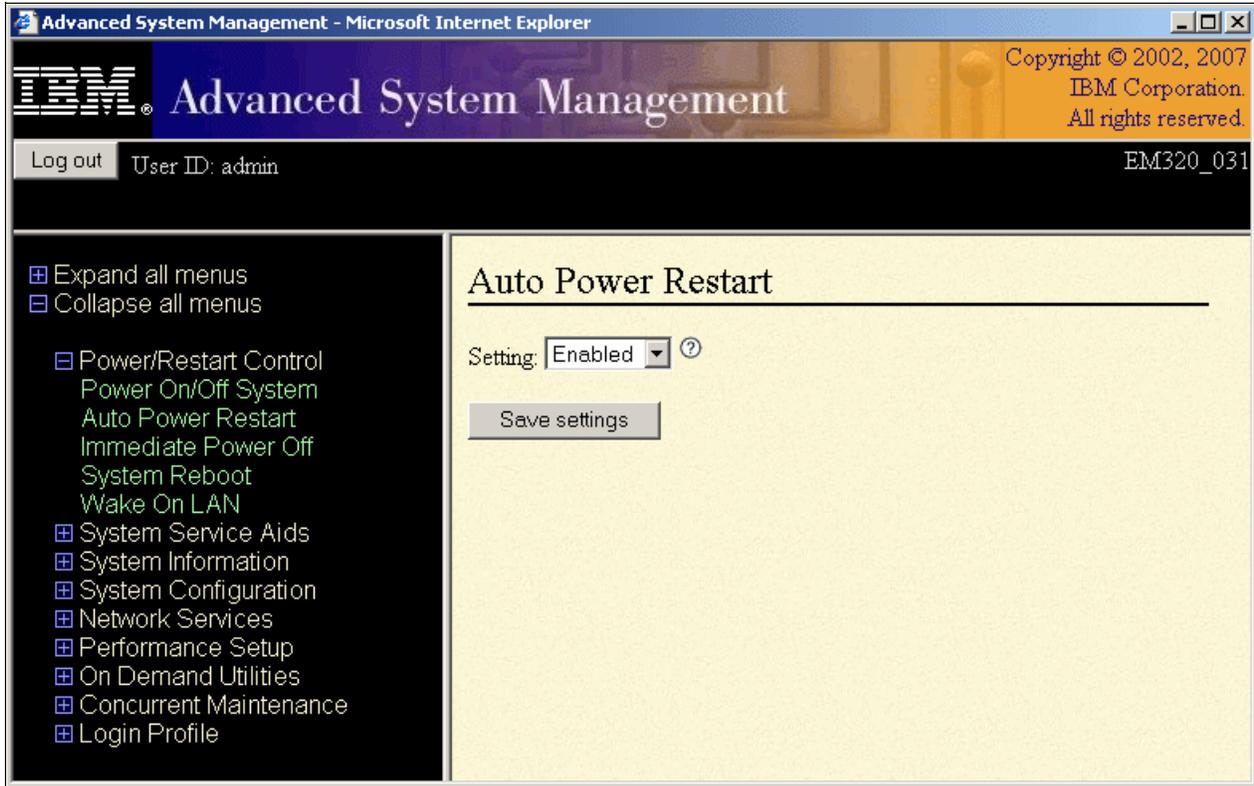
Figure 4-2   ASMI Auto Power Restart setting

Dynamic processor deallocation

Dynamic processor deallocation enables automatic deconfiguration of processor cores when patterns of recoverable errors, for example correctable errors on processor caches, are detected. Dynamic processor deallocation prevents a recoverable error from escalating to an unrecoverable system error, which might otherwise result in an unscheduled server outage. Dynamic processor deallocation relies on the service processor's ability to use FFDC-generated recoverable error information to notify the POWER Hypervisor when a processor core reaches its predefined error limit. Then the POWER Hypervisor, in conjunction with the operating system, redistributes the work to the remaining processor cores, deallocates the offending processor core, and continues normal operation. As necessary, this can revert from simultaneous multiprocessing to uniprocessor processing.

IBM's logical partitioning strategy allows any processor core to be shared with any partition on the system up to the processor capacity represented by the number of physically activated processor cores.

The capacity on demand capabilities affect what happens when a processor core is determined to be non-operational. The capacity on demand capabilities supported by the Power 520 8203-E4A and Power 550 8204-E8A models are restricted to permanent processor activation (CoD) only. That is, a one processor unit capacity must be permanently activated for that capacity to be available for system use as controlled by the POWER Hypervisor.

The Power 520 9408-M25 can be a 1-2 Way configuration with the second processor unactivated. The Power 550 9409-M50 can be a 1-4 Way configuration with the second through fourth processors unactivated.

If you have a permanent processor error and there are available unactivated processor cores, the system will effectively put the unactivated processor into operation once it has been determined an activated processor is no longer operational. This means, you still have the processor power you are licensed for when the failed processor is "deallocated" and an unactivated processor takes its place.

4.2.3 Processor instruction retry

Soft failures in the processor core are transient errors. When an error is encountered in the core, the POWER6 processor will first automatically retry the instruction. If the source of the error was truly transient, the instruction will succeed and the system will continue as before.

On predecessor IBM systems, this error would have caused a checkstop.

4.2.4 Processor checkstop

Another major advancement in POWER6 processors is single processor checkstopping. A processor checkstop would result in a system checkstop. A new feature available with POWER6 is the ability to contain most processor checkstops to the partition that was using the processor at the time full checkstop goes into effect. This significantly reduces the probability of any one processor affecting total system availability.

4.2.5 Alternate processor retry

Hard processor failures are more difficult, being true logical errors that will be replicated each time the instruction is repeated. Retrying the instruction does not help in this situation because the instruction will continue to fail.

Systems with POWER6 processors introduce the ability to extract the failing instruction from the faulty core and retry it elsewhere in the system, after which the failing core is dynamically deconfigured and called out for replacement.

The entire process is transparent to the partition owning the failing instruction. Systems with POWER6 processors are designed to avoid what would have been a full system.

When a full processor core has been deconfigured, either a previously unactivated processor core has taken its place or, if that is not possible, system-wide total processor capacity has been lowered below the licensed number.

It is time to contact your service representative and perhaps make workload or partition definition changes until the problem has been resolved.

4.2.6 Memory protection

Memory and cache arrays comprise data bit lines that feed into a memory word. A memory word is addressed by the system as a single element. Depending on the size and

addressability of the memory element, each data bit line might include thousands of individual bits or memory cells. For example:

► A single memory module on a Dual Inline Memory Module (DIMM) can have a capacity of 1 GB, and supply eight bit lines of data for an ECC word. In this case, each bit line in the ECC word holds 128 MB behind it, corresponding to more than 128 million memory cell addresses.

► A 32 KB L1 cache with a 16-byte memory word, on the other hand, would have only 2 KB behind each memory bit line.

A memory protection architecture that provides good error resilience for a relatively small L1 cache might be very inadequate for protecting the much larger system main store. Therefore, a variety of different protection methods are used in POWER6 processor-based systems to avoid uncorrectable errors in memory.

Memory protection plans must take into account many factors, including:

► Size

► Desired performance

► Memory array manufacturing characteristics.

POWER6 processor-based systems have a number of protection schemes designed to prevent, protect, or limit the effect of errors in main memory. These capabilities include:

Hardware scrubbing   Hardware scrubbing is a method used to deal with soft errors. IBM POWER6 processor-based systems periodically address all memory locations and any memory locations with an ECC error are rewritten with the correct data.

Error correcting code   Error correcting code (ECC) allows a system to detect up to two errors in a memory word and correct one of them. However, without additional correction techniques, if more than one bit is corrupted, a system will fail.

Chipkill   Chipkill is an enhancement to ECC that enables a system to sustain the failure of an entire DRAM. Chipkill spreads the bit lines from a DRAM over multiple ECC words, so that a catastrophic DRAM failure would affect at most one bit in each word. Barring a future single bit error, the system can continue indefinitely in this state with no performance degradation until the failed DIMM can be replaced.

Redundant bit steering IBM systems use redundant bit steering to avoid situations where multiple single-bit errors align to create a multi-bit error. In the event that an IBM POWER6 processor-based system detects an abnormal number of errors on a bit line, it can dynamically steer the data stored at this bit line into one of a number of spare lines.
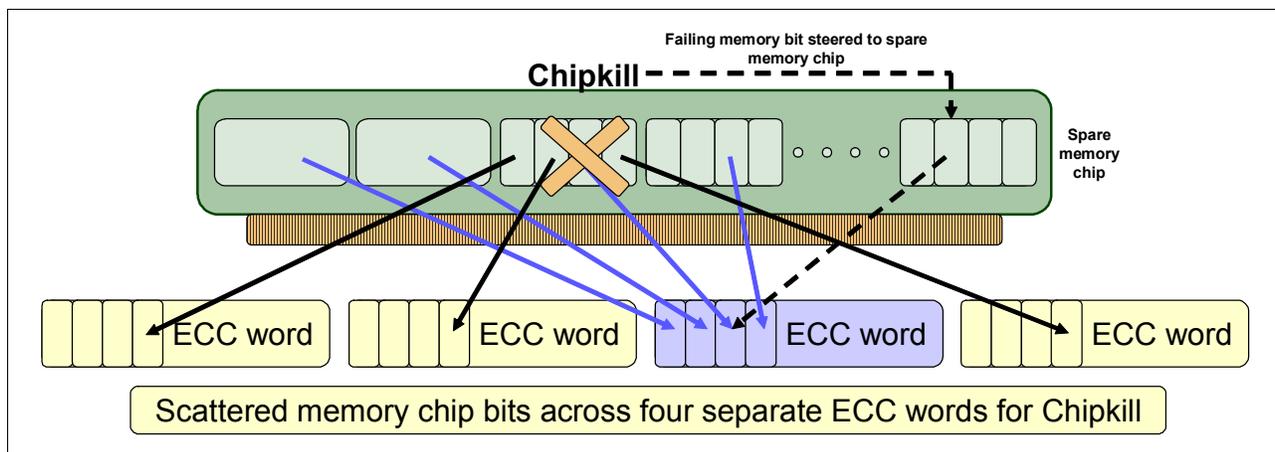
Figure 4-3   Memory protection capabilities in action

Memory page deallocation

While coincident single cell errors in separate memory chips is a statistic rarity, POWER6 processor-based systems can contain these errors using a memory page deallocation scheme for partitions running AIX and for memory pages owned by the POWER Hypervisor. If a memory address experiences an uncorrectable or repeated correctable single cell error, the Service Processor sends the memory page address to the POWER Hypervisor to be marked for deallocation.

The operating system performs memory page deallocation without any user intervention and is transparent to end users and applications.

The POWER Hypervisor maintains a list of pages marked for deallocation during the current platform IPL. During a partition IPL, the partition receives a list of all the bad pages in its address space.

In addition, if memory is dynamically added to a partition (through a Dynamic LPAR operation), the POWER Hypervisor warns the operating system if memory pages are included which need to be deallocated.

Finally, should an uncorrectable error occur, the system can deallocate the memory group associated with the error on all subsequent system reboots until the memory is repaired. This is intended to guard against future uncorrectable errors while waiting for parts replacement.

> **Note:** Memory page deallocation handles single cell failures, but, because of the sheer size of data in a data bit line, it may be inadequate for dealing with more catastrophic failures. Redundant bit steering will continue to be the preferred method for dealing with these types of problems.

Memory control hierarchy

A memory controller on a POWER6 processor-based system is designed with four ports. Each port connects up to three DIMMS using a daisy-chained bus. The memory bus supports ECC checking on data, addresses, and command information. A spare line on the bus is also available for repair using a self-healing strategy. In addition, ECC checking on addresses and commands is extended to DIMMs on DRAMs. Because it uses a daisy-chained memory access topology, this system can deconfigure a DIMM that encounters a DRAM fault, without deconfiguring the bus controller, even if the bus controller is contained on the DIMM.
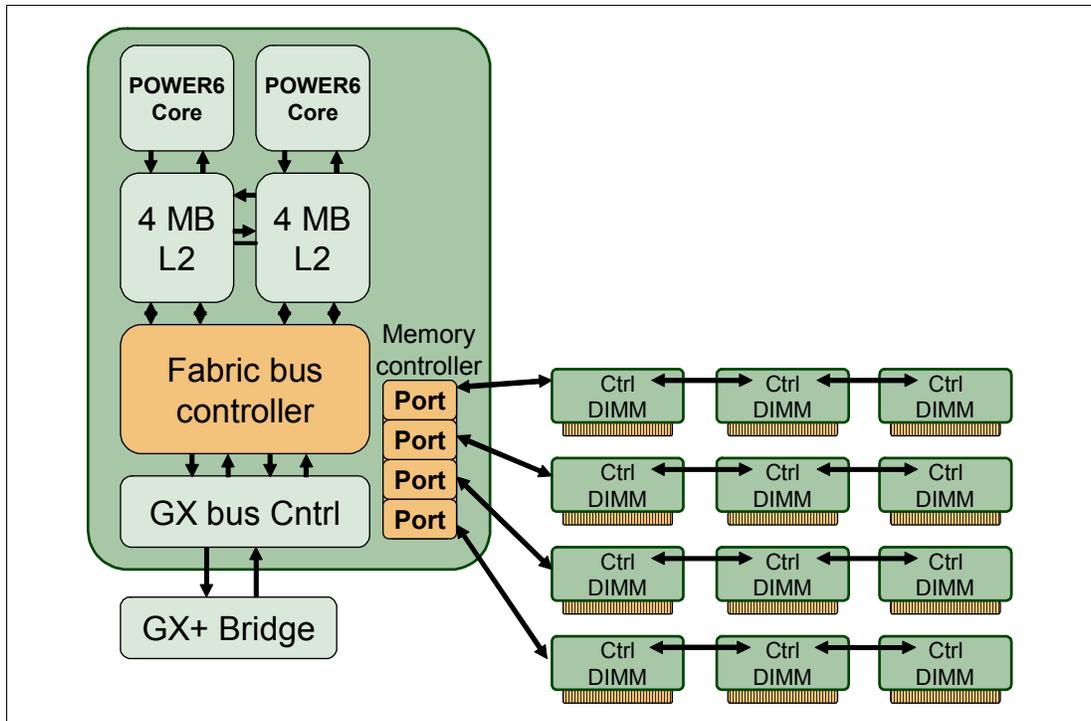
Figure 4-4   Memory control hierarchy

Memory deconfiguration

Defective memory discovered at boot time is automatically switched off, unless it is already the minimum amount required to boot. If the Service Processor detects a memory fault at boot time, it marks the affected memory as bad so it is not to be used on subsequent reboots (Memory Persistent Deallocation).

If the Service Processor identifies faulty memory in a server that includes CoD memory, the POWER Hypervisor attempts to replace the faulty memory with available CoD memory. Faulty resources are marked as deallocated and working resources are included in the active memory space. Since these activities reduce the amount of CoD memory available for future use, repair of the faulty memory should be scheduled as soon as is convenient.

Upon reboot, if not enough memory is available to meet minimum partition requirements the POWER Hypervisor will reduce the capacity of one or more partitions. The HMC receives notification of the failed component, triggering a service call.

4.2.7 Special uncorrectable error handling

While it is rare, an uncorrectable data error can occur in memory or a cache. POWER6 processor-based systems attempt to limit, to the least possible disruption, the impact of an uncorrectable error using a well-defined strategy that first considers the data source. Sometimes, an uncorrectable error is temporary in nature and occurs in data that can be recovered from another repository. For example:

► Data in the instruction L1 cache is never modified within the cache itself. Therefore, an uncorrectable error discovered in the cache is treated like an ordinary cache miss, and correct data is loaded from the L2 cache.

In cases where the data cannot be recovered from another source, a technique called Special Uncorrectable Error (SUE) handling is used to determine whether the corruption is truly a

threat to the system. If, as may sometimes be the case, the data is never actually used but is simply over-written, then the error condition can safely be voided and the system will continue to operate normally.

When an uncorrectable error is detected, the system modifies the associated ECC word, thereby signaling to the rest of the system that the "standard" ECC is no longer valid. The Service Processor is then notified, and takes appropriate actions. When running AIX V5.2 or greater or Linux[1] and a process attempts to use the data, the OS is informed of the error and terminates only the specific user program.

It is only in the case where the corrupt data is used by the POWER Hypervisor that the entire system must be rebooted, thereby preserving overall system integrity.

Depending on system configuration and source of the data, errors encountered during I/O operations may not result in a machine check. Instead, the incorrect data is handled by the processor host bridge (PHB) chip. When the PHB chip detects a problem it rejects the data, preventing data being written to the I/O device. The PHB then enters a freeze mode halting normal operations. Depending on the model and type of I/O being used, the freeze may include the entire PHB chip, or simply a single bridge. This results in the loss of all I/O operations that use the frozen hardware until a power-on reset of the PHB. The impact to partitions depends on how the I/O is configured for redundancy. In a server configured for fail-over availability, redundant adapters spanning multiple PHB chips could enable the system to recover transparently, without partition loss.

4.2.8 Cache protection mechanisms

POWER6 processor-based systems are designed with cache protection mechanisms, including cache line delete in L2 arrays, Processor Instruction Retry and Alternate Processor Recovery protection on L1-I and L1-D, and redundant "Repair" bits in L1-I, L1-D, and L2 caches, as well as L2 directories.

L1 instruction and data array protection
The POWER6 processor's instruction and data caches are protected against temporary errors using the POWER6 Processor Instruction Retry feature and against solid failures by Alternate Processor Recovery, both mentioned earlier. In addition, faults in the SLB array are recoverable by the POWER Hypervisor.

L2 Array Protection
On a POWER6 processor-based system, the L2 cache is protected by ECC, which provides single-bit error correction and double-bit error detection. Single-bit errors are corrected before forwarding to the processor, and subsequently written back to L2. Like the other data caches and main memory, uncorrectable errors are handled during run-time by the Special Uncorrectable Error handling mechanism. Correctable cache errors are logged and if the error reaches a threshold, a Dynamic Processor Deallocation event is initiated.

Starting with POWER6 processor-based systems, the L2 cache is further protected by incorporating a dynamic cache line delete algorithm. Up to six L2 cache lines can be automatically deleted. It is not likely that deletion of a few cache lines will adversely affect server performance. When six cache lines have been repaired, the L2 is marked for persistent deconfiguration on subsequent system reboots until it can be replaced.

---

[1] SLES 10 SP1 or later, and in RHEL 4.5 or later (including RHEL 5.1).

4.2.9 PCI error recovery

IBM estimates that PCI adapters can account for a significant portion of the hardware-based errors on a large server. While servers that rely on boot-time diagnostics can identify failing components to be replaced by hot-swap and reconfiguration, runtime errors pose a more significant problem.

PCI adapters are generally complex designs involving extensive on-board instruction processing, often on embedded microcontrollers. They tend to use industry standard grade components with an emphasis on product cost relative to high reliability. In some cases they may be more likely to encounter internal microcode errors, or many of the hardware errors described for the rest of the server.

The traditional means of handling these problems is through adapter internal error reporting and recovery techniques in combination with operating system device driver management and diagnostics. In some cases, an error in the adapter may cause transmission of bad data on the PCI bus itself, resulting in a hardware detected parity error and causing a global machine check interrupt, eventually requiring a system reboot to continue.

In 2001, IBM introduced a methodology that uses a combination of system firmware and Extended Error Handling (EEH) device drivers that allows recovery from intermittent PCI bus errors. This approach works by recovering and resetting the adapter, thereby initiating system recovery for a permanent PCI bus error. Rather than failing immediately, the faulty device is frozen and restarted, preventing a machine check. POWER6 technology extends this capability to PCIe bus errors, and includes expanded Linux support for EEH as well.
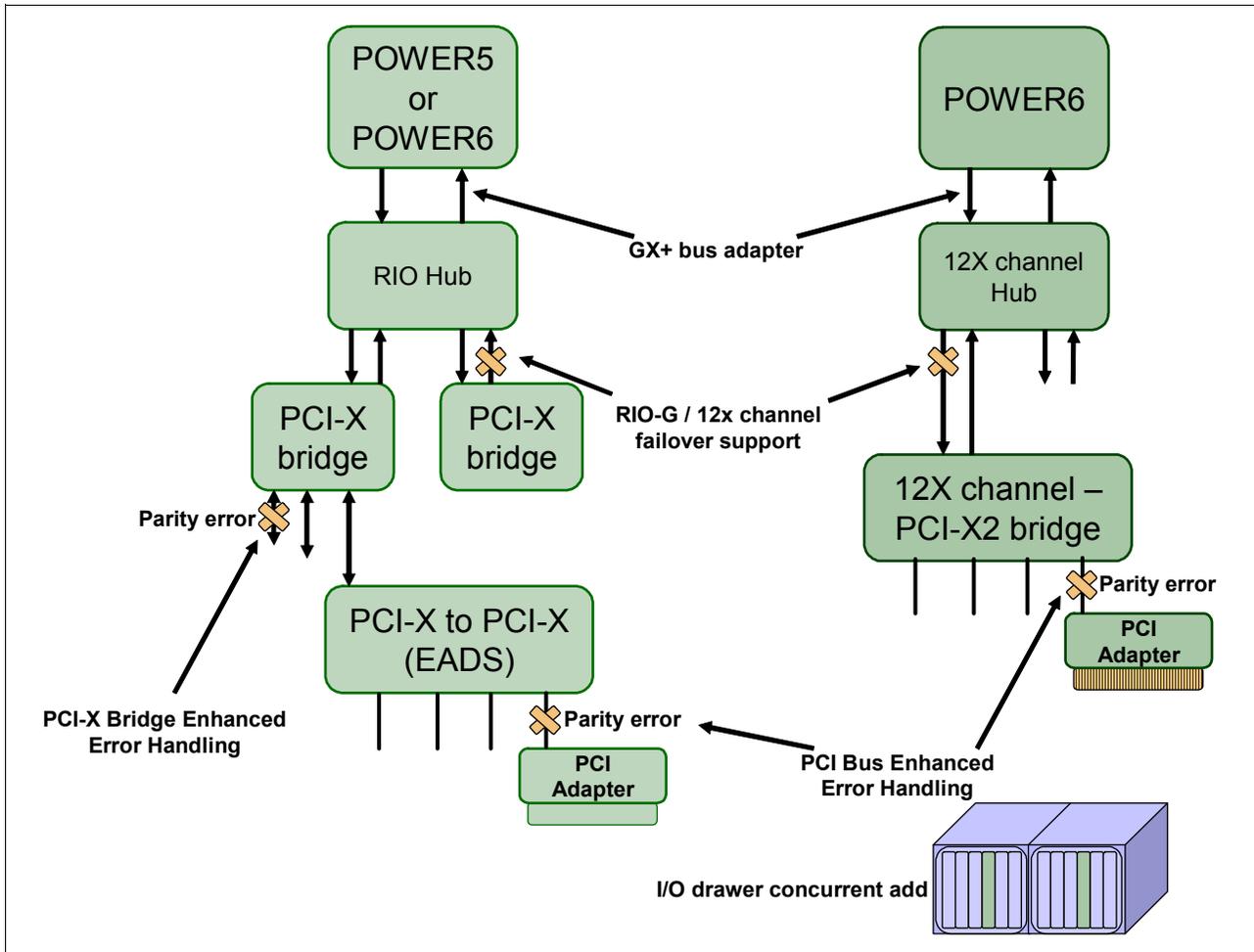
Figure 4-5   PCI error recovery

4.3 Serviceability

The IBM POWER6 Serviceability strategy evolves from, and improves upon, the service architecture deployed on the POWER5 processor-based systems. The IBM service team has enhanced the base service capabilities and continues to implement a strategy that incorporates best-of-breed service characteristics from IBM's diverse System x, System i, System p, and high-end System z® offerings.

The goal of the IBM Serviceability Team is to design and provide the most efficient system service environment that incorporates:

► Easy access to service components

► On demand service education

► An automated guided repair strategy that uses common service interfaces for a converged service approach across multiple IBM server platforms

By delivering on these goals, POWER6 processor-based systems enable faster and more accurate repair while reducing the possibility of human error.

Client control of the service environment extends to firmware maintenance on all of the POWER6 processor-based systems, including the Power 520. This strategy contributes to higher systems availability with reduced maintenance costs.

This section provides an overview of the progressive steps of error detection, analysis, reporting, and repairing found in all POWER6 processor-based systems.

4.3.1 Detecting errors

The first and most crucial component of a solid serviceability strategy is the ability to accurately and effectively detect errors when they occur. While not all errors are a guaranteed threat to system availability, those that go undetected can cause problems because the system does not have the opportunity to evaluate and act if necessary. POWER6 processor-based systems employ System z server inspired error detection mechanisms that extend from processor cores and memory to power supplies and hard drives.

Service Processor

The Service Processor is a separately powered microprocessor, separate from the main instruction processing complex. The Service Processor enables POWER Hypervisor and Hardware Management Console surveillance, selected remote power control, environmental monitoring, reset and boot features, remote maintenance and diagnostic activities, including console mirroring. On systems without a Hardware Management Console, the Service Processor can place calls to report surveillance failures with the POWER Hypervisor, critical environmental faults, and critical processing faults even when the main processing unit is inoperable. The Service Processor provides services common to modern computers such as:

► Environmental monitoring

– The Service Processor monitors the server's built-in temperature sensors, sending instructions to the system fans to increase rotational speed when the ambient temperature is above the normal operating range.

– Using an architected operating system interface, the Service Processor notifies the operating system of potential environmental related problems (for example, air conditioning and air circulation around the system) so that the system administrator can take appropriate corrective actions before a critical failure threshold is reached.

– The Service Processor can also post a warning and initiate an orderly system shutdown for a variety of other conditions:

• When the operating temperature exceeds the critical level (for example, failure of air conditioning or air circulation around the system).

• When the system fan speed is out of operational specification, for example, due to a fan failure, the system can increase speed on the redundant fans to compensate this failure or take other actions.

• When the server input voltages are out of operational specification.

► Mutual Surveillance

The Service Processor monitors the operation of the POWER Hypervisor firmware during the boot process and watches for loss of control during system operation. It also allows the POWER Hypervisor to monitor Service Processor activity. The Service Processor can take appropriate action, including calling for service, when it detects the POWER Hypervisor firmware has lost control. Likewise, the POWER Hypervisor can request a Service Processor repair action if necessary.

- ► Availability

   The auto-restart (reboot) option, when enabled, can reboot the system automatically following an unrecoverable firmware error, firmware hang, hardware failure, or environmentally induced (AC power) failure.

- ► Fault Monitoring

   BIST (built-in self-test) checks processor, memory, and associated hardware required for proper booting of the operating system, when the system is powered on at the initial install or after a hardware configuration change (for example, an upgrade). If a non-critical error is detected or if the error occurs in a resource that can be removed from the system configuration, the booting process is designed to proceed to completion. The errors are logged in the system nonvolatile random access memory (NVRAM). When the operating system completes booting, the information is passed from the NVRAM into the system error log where it is analyzed by error log analysis (ELA) routines. Appropriate actions are taken to report the boot time error for subsequent service if required.

One important Service Processor improvement allows the system administrator or service representative dynamic access to the Advanced Systems Management Interface (ASMI) menus. In previous generations of servers, these menus were only accessible when the system was in standby power mode. Now, the menus are available from any Web browser-enabled console attached to the Ethernet service network concurrent with normal system operation. A user with the proper access authority and credentials can now dynamically modify service defaults, interrogate Service Processor progress and error logs, set and reset guiding light LEDs – indeed, access all Service Processor functions without having to power-down the system to the standby state.

The Service Processor also manages the interfaces for connecting Uninterruptible Power Source (UPS) systems to the POWER6 processor-based systems, performing Timed Power-On (TPO) sequences, and interfacing with the power and cooling subsystem.

Error checkers

IBM POWER6 processor-based systems contain specialized hardware detection circuitry that is used to detect erroneous hardware operations. Error checking hardware ranges from parity error detection coupled with processor instruction retry and bus retry, to ECC correction on caches and system buses. All IBM hardware error checkers have distinct attributes:

- ► Continually monitoring system operations to detect potential calculation errors.
- ► Attempting to isolate physical faults based on run time detection of each unique failure.
- ► Ability to initiate a wide variety of recovery mechanisms designed to correct the problem. The POWER6 processor-based systems include extensive hardware and firmware recovery logic.

Fault Isolation Registers

Error checker signals are captured and stored in hardware Fault Isolation Registers (FIRs). The associated "Who's on First" logic circuitry is used to limit the domain of an error to the first checker that encounters the error. In this way, run-time error diagnostics can be deterministic such that for every check station, the unique error domain for that checker is defined and documented. Ultimately, the error domain becomes the Field Replaceable Unit (FRU) call, and manual interpretation of the data is not normally required.

First Failure Data Capture (FFDC)

First Failure Data Capture (FFDC) is an error isolation technique that ensures that when a fault is detected in a system through error checkers or other types of detection methods, the

root cause of the fault will be captured without the need to recreate the problem or run an extended tracing or diagnostics program.

For the vast majority of faults, a good FFDC design means that the root cause will be detected automatically without intervention by a service representative. Pertinent error data related to the fault is captured and saved for analysis. In hardware, FFDC data is collected from the fault isolation registers and "Who's on First" logic. In firmware, this data consists of return codes, function calls, and so forth.

FFDC "check stations" are carefully positioned within the server logic and data paths to ensure that potential errors can be quickly identified and accurately tracked to a Field Replaceable Unit (FRU).

This proactive diagnostic strategy is a significant improvement over the classic, less accurate "reboot and diagnose" service approaches.

Figure 4-6 shows a schematic of a Fault Isolation Register implementation.
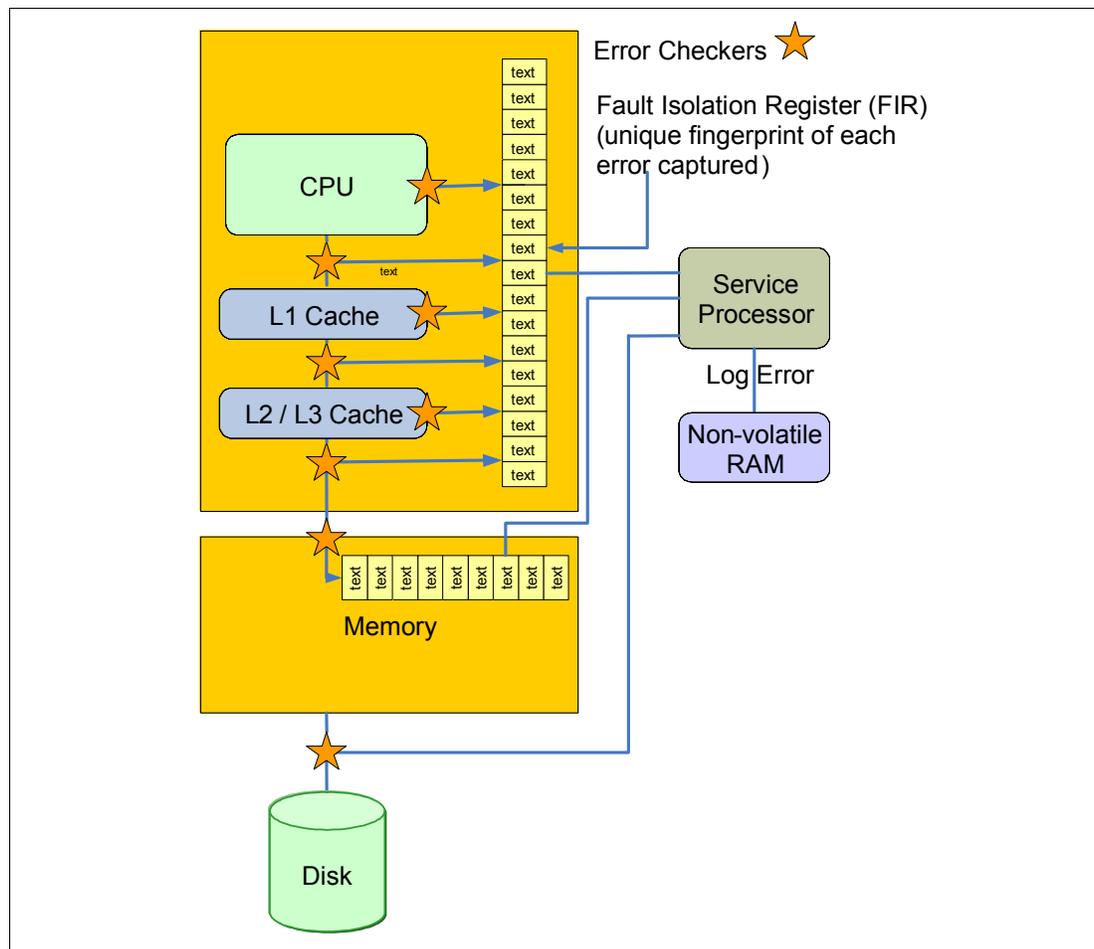
Figure 4-6   Schematic of an FIR implementation

Fault isolation

The Service Processor interprets error data captured by the FFDC checkers—saved in the FIRs and 'Who's On First logic or other firmware related data capture methods—in order to determine the root cause of the error event.

Root cause analysis might indicate that the event is recoverable, meaning that a service action point or need for repair has not been reached. Alternatively, it could indicate that a service action point has been reached, where the event exceeded a pre-determined threshold or was unrecoverable. Based upon the isolation analysis, recoverable error threshold counts may be incremented. No specific service action could be necessary when the event is recoverable.

When the event requires a service action, additional required information will be collected to service the fault. For unrecoverable errors or for recoverable events that meet or exceed their service threshold—meaning a service action point has been reached—a request for service will be initiated through an error logging component.

4.3.2 Diagnosing problems

Using the extensive network of advanced and complementary error detection logic built directly into hardware, firmware, and operating systems, the IBM POWER6 processor-based Systems can perform considerable self diagnosis.

Boot time

When an IBM POWER6 processor-based system powers up, the Service Processor initializes system hardware. Boot-time diagnostic testing uses a multi-tier approach for system validation, starting with managed low level diagnostics supplemented with system firmware initialization and configuration of I/O hardware, followed by OS-initiated software test routines. Boot-time diagnostic routines include:

► Built-in-Self-Tests (BISTs) for both logic components and arrays ensure the internal integrity of components. Because the Service Processor assists in performing these tests, the system is enabled to perform fault determination and isolation whether system processors are operational or not. Boot-time BISTs may also find faults undetectable by processor-based Power-on-Self-Test (POST) or diagnostics.

► Wire Tests discover and precisely identify connection faults between components such as processors, memory, or I/O hub chips.

► Initialization of components such as ECC memory, typically by writing patterns of data and allowing the server to store valid ECC data for each location, can help isolate errors.

To minimize boot time, the system will determine which of the diagnostics are required to be started in order to ensure correct operation based on the way the system was powered off, or on the boot-time selection menu.

Run time

All POWER6 processor-based systems can monitor critical system components during run time, and they can take corrective actions when recoverable faults occur. IBM's hardware error check architecture provides the ability to report non-critical errors in an "out-of-band" communications path to the Service Processor without affecting system performance.

A significant part of IBM's runtime diagnostic capabilities originate with the POWER6 Service Processor. Extensive diagnostic and fault analysis routines have been developed and improved over many generations of POWER processor-based servers, and enable quick and accurate predefined responses to both actual and potential system problems.

The Service Processor correlates and processes runtime error information, using logic derived from IBM's engineering expertise to count recoverable errors (called *Thresholding*)

and predict when corrective actions must be automatically initiated by the system. These actions can include:

► Requests for a part to be replaced

► Dynamic (on-line) invocation of built-in redundancy for automatic replacement of a failing part

► Dynamic deallocation of failing components so that system availability is maintained

Device drivers

In certain cases diagnostics are best performed by operating system-specific drivers, most notably I/O devices that are owned directly by a logical partition. In these cases, the operating system device driver will often work in conjunction with I/O device microcode to isolate and recover from problems. Potential problems are reported to an operating system device driver, which logs the error. I/O devices can also include specific exercisers that can be invoked by the diagnostic facilities for problem recreation if required by service procedures.

4.3.3 Reporting problems

In the unlikely event that a system hardware or environmentally induced failure is diagnosed, POWER6 processor-based systems report the error through a number of mechanisms. This ensures that appropriate entities are aware that the system may be operating in an error state. However, a crucial piece of a solid reporting strategy is ensuring that a single error communicated through multiple error paths is correctly aggregated, so that later notifications are not accidently duplicated.

Error logging and analysis

Once the root cause of an error has been identified by a fault isolation component, an error log entry is created with some basic data such as:

► An error code uniquely describing the error event

► The location of the failing component

► The part number of the component to be replaced, including pertinent data like engineering and manufacturing levels

► Return codes

► Resource identifiers

► First Failure Data Capture data

Data containing information on the effect that the repair will have on the system is also included. Error log routines in the operating system can then use this information and decide to call home to contact service and support, send a notification message, or continue without an alert.

Remote support

The Remote Management and Control (RMC) application is delivered as part of the base operating system, including the operating system running on the Hardware Management Console. RMC provides a secure transport mechanism across the LAN interface between the operating system and the Hardware Management Console and is used by the operating system diagnostic application for transmitting error information. It performs a number of other functions as well, but these are not used for the service infrastructure.

Manage serviceable events

A critical requirement in a logically partitioned environment is to ensure that errors are not lost before being reported for service, and that an error should only be reported once, regardless of how many logical partitions experience the potential effect of the error. The "Manage Serviceable Events" task on the Hardware Management Console (HMC) is responsible for aggregating duplicate error reports, and ensures that all errors are recorded for review and management.

When a local or globally reported service request is made to the operating system, the operating system diagnostic subsystem uses the Remote Management and Control Subsystem (RMC) to relay error information to the Hardware Management Console. For global events (platform unrecoverable errors, for example) the Service Processor will also forward error notification of these events to the Hardware Management Console, providing a redundant error-reporting path in case of errors in the RMC network.

The first occurrence of each failure type will be recorded in the *Manage Serviceable Events* task on the Hardware Management Console. This task will then filter and maintain a history of duplicate reports from other logical partitions or the Service Processor. It then looks across all active service event requests, analyzes the failure to ascertain the root cause and, if enabled, initiates a call home for service. This methodology ensures that all platform errors will be reported through at least one functional path, ultimately resulting in a single notification for a single problem.

Extended error data (EED)

Extended error data (EED) is additional data that is collected either automatically at the time of a failure or manually at a later time. The data collected is dependent on the invocation method but includes information like firmware levels, operating system levels, additional fault isolation register values, recoverable error threshold register values, system status, and any other pertinent data.

The data is formatted and prepared for transmission back to IBM to assist the service support organization with preparing a service action plan for the service representative or for additional analysis.

System dump handling

In some circumstances, an error may require a dump to be automatically or manually created. In this event, it will be offloaded to the HMC upon reboot. Specific HMC information is included as part of the information that can optionally be sent to IBM support for analysis. If additional information relating to the dump is required, or if it becomes necessary to view the dump remotely, the HMC dump record will notify IBM's support center regarding which HMC the dump is located on.

4.3.4 Notifying the appropriate contacts

Once a POWER6 processor-based system has detected, diagnosed, and reported an error to an appropriate aggregation point, it then takes steps to notify the client, and if necessary the IBM Support Organization. Depending upon the assessed severity of the error and support agreement, this could range from a simple notification to having field service personnel automatically dispatched to the client site with the correct replacement part.

Customer notify

When an event is important enough to report, but does not indicate the need for a repair action or the need to call home to IBM service and support, it is classified as *customer notify*. Customers are notified because these events might be of interest to an administrator. The

event might be a symptom of an expected systemic change, such as a network reconfiguration or failover testing of redundant power or cooling systems. Examples of these events include:

- ► Network events like the loss of contact over a Local Area Network (LAN)
- ► Environmental events such as ambient temperature warnings
- ► Events that need further examination by the client, but these events do not necessarily require a part replacement or repair action

Customer notify events are serviceable events by definition because they indicate that something has happened which requires client awareness in the event they want to take further action. These events can always be reported back to IBM at the client's discretion.

Call home

A correctly configured POWER6 processor-based system can initiate an automatic or manual call from a client location to the IBM service and support organization with error data, server status, or other service-related information. *Call home* invokes the service organization in order for the appropriate service action to begin, automatically opening a problem report and in some cases also dispatching field support. This automated reporting provides faster and potentially more accurate transmittal of error information. While configuring call home is optional, clients are strongly encouraged to configure this feature in order to obtain the full value of IBM service enhancements.

Vital product data (VPD) and inventory management

POWER6 processor-based systems store vital product data (VPD) internally, which keeps a record of how much memory is installed, how many processors are installed, manufacturing level of the parts, and so on. These records provide valuable information that can be used by remote support and service representatives, enabling them to provide assistance in keeping the firmware and software on the server up-to-date.

IBM problem management database

At the IBM support center, historical problem data is entered into the IBM Service and Support Problem Management database. All of the information related to the error along with any service actions taken by the service representative are recorded for problem management by the support and development organizations. The problem is then tracked and monitored until the system fault is repaired.

4.3.5 Locating and repairing the problem

The final component of a comprehensive design for serviceability is the ability to effectively locate and replace parts requiring service. POWER6 processor-based systems utilize a combination of visual cues and guided maintenance procedures to ensure that the identified part is replaced correctly, every time.

Guiding Light LEDs

Guiding Light uses a series of flashing LEDs, allowing a service provider to quickly and easily identify the location of system components. Guiding Light can also handle multiple error conditions simultaneously, which could be necessary in some very complex high-end configurations.

In the Guiding Light LED implementation, when a fault condition is detected on a POWER6 processor-based system, an amber System Attention LED will be illuminated. Upon arrival, the service provider engages *identify* mode by selecting a specific problem. The Guiding

Light system then identifies the part that needs to be replaced by flashing the amber *identify* LED.

Datacenters can be complex places, and Guiding Light is designed to do more than identify visible components. When a component might be hidden from view, Guiding Light can flash a sequence of LEDs that extend to the frame exterior, clearly "guiding" the service representative to the correct rack, system, enclosure, drawer, and component.

The control (operator) panel

The control panel on a POWER6 processor-based system is a 4-row by 16-element LCD display used to present boot progress codes, indicating advancement through the system power-on and initialization processes. The Operator Panel is also used to display error and location codes when an error occurs that prevents the system from booting. The operator panel includes several buttons allowing a service representative or the client to change various boot-time options, and perform a subset of the service functions that are available on the Advanced System Management Interface (ASMI).

See 2.15.2, "Control (operator) panel" on page 122 for additional control panel information.

Concurrent maintenance

The IBM POWER6 processor-based systems are designed with the understanding that certain components have higher intrinsic failure rates than others. The movement of fans, power supplies, and physical storage devices naturally make them more susceptible to wearing down or burning out, while other devices such as I/O adapters may begin to wear from repeated plugging and unplugging. For this reason, these devices are specifically designed to be concurrently maintainable, when properly configured.

In other cases, a client may be in the process of moving or redesigning a datacenter, or planning a major upgrade. At times like these, flexibility is crucial. The IBM POWER6 processor-based systems are designed for redundant or concurrently maintainable power, fans, physical storage, and I/O towers.

Blind-swap PCI adapters

Blind-swap PCI adapters represent significant service and ease-of-use enhancements in I/O subsystem design while maintaining high PCI adapter density.

Standard PCI designs supporting *hot-add* and *hot-replace* require top access so that adapters can be slid into the PCI I/O slots vertically. *Blind-swap* allows PCI adapters to be concurrently replaced without having to put the I/O drawer into a service position.

Firmware updates

Firmware updates for POWER6 processor-based systems are released in a cumulative sequential fix format, packaged as an RPM for concurrent application and activation. Administrators can install and activate many firmware patches without cycling power or rebooting the server.

The new firmware image is loaded on the HMC using any of the following methods:

► IBM-distributed media such as a CD-ROM

► A Problem Fix distribution from the IBM Service and Support repository

► Download from the IBM Web site:

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

For additional information in this publication, see 4.5.6, "IBM Power Systems firmware maintenance" on page 188.

► FTP from another server

IBM will support multiple firmware releases in the field, so under expected circumstances a server can operate on an existing firmware release, using concurrent firmware fixes to stay up-to-date with the current patch level. Because changes to some server functions (for example, changing initialization values for chip controls) cannot occur during system operation, a patch in this area will require a system reboot for activation. Under normal operating conditions, IBM intends to provide patches for an individual firmware release level for up to two years after first making the release code generally availability. After this period, clients should plan to update in order to stay on a supported firmware release.

Activation of new firmware functions, as opposed to patches, will require installation of a new firmware release level. This process is disruptive to server operations in that it requires a scheduled outage and full server reboot.

In addition to concurrent and disruptive firmware updates, IBM will also offer concurrent patches that include functions which are not activated until a subsequent server reboot. A server with these patches will operate normally. The additional concurrent fixes will be installed and activated when the system reboots after the next scheduled outage.

Additional capability is being added to the POWER6 firmware to be able to view the status of a system power control network background firmware update. This subsystem will update as necessary as migrated nodes or I/O drawers are added to the configuration. The new firmware will not only provide an interface to be able to view the progress of the update, but also control starting and stopping of the background update if a more convenient time becomes available.

Repair and verify

Repair and verify (R&V) is a system used to guide a service provider step-by-step through the process of repairing a system and verifying that the problem has been repaired. The steps are customized in the appropriate sequence for the particular repair for the specific system being repaired. Repair scenarios covered by repair and verify include:

► Replacing a defective Field Replaceable Unit (FRU)

► Reattaching a loose or disconnected component

► Correcting a configuration error

► Removing or replacing an incompatible FRU

► Updating firmware, device drivers, operating systems, middleware components, and IBM applications after replacing a part

► Installing a new part

Repair and verify procedures are designed to be used both by service representative providers who are familiar with the task at hand and those who are not. On Demand Education content is placed in the procedure at the appropriate locations. Throughout the repair and verify procedure, repair history is collected and provided to the Service and Support Problem Management Database for storage with the Serviceable Event, to ensure that the guided maintenance procedures are operating correctly.

Clients can subscribe through the Subscription Services to obtain the notifications on the latest updates available for service-related documentation. The latest version of the documentation is accessible through the Internet, and a CD-ROM-based version is also available.

4.4 Operating system support for RAS features

Table 4-1 gives an overview of a number of features for continuous availability supported by the different operating systems running on the POWER6 processor-based systems.

Table 4-1   Operating system support for selected RAS features

RAS feature	AIX V5.3	AIX V6.1	IBM i 5.4 LIC 5.4.5	IBM i 6.1	RHEL V5.1	SLES V10
System Deallocation of Failing Components						
Dynamic processor deallocation	Y	Y	Y	Y	Y[1]	Y
Dynamic processor sparing	Y	Y	Y	Y	Y	Y
Processor instruction retry	Y	Y	Y	Y	Y	Y
Alternate processor recovery	Y	Y	Y	Y	Y	Y
Partition contained checkstop	Y	Y	Y	Y	Y	Y
Persistent processor deallocation	Y	Y	Y	Y	Y	Y
GX+ bus persistent deallocation	Y	Y	Y	Y	N	N
PCI bus extended error detection	Y	Y	Y	Y	Y	Y
PCI bus extended error recovery	Y	Y	Y	Y	Limited[1]	Limited
PCI-PCI bridge extended error handling	Y	Y	Y	Y	N	N
Redundant RIO Link	Y	Y	Y	Y	Y	Y
PCI card hot swap	Y	Y	Y	Y	Y[1]	Y
Dynamic SP failover at runtime	Y	Y	Y	Y	N	N
Clock failover at IPL	Y	Y	Y	Y	Y	Y
Memory Availability						
ECC Memory, L2 cache	Y	Y	Y	Y	Y	Y
Dynamic bit-steering (spare memory)	Y	Y	Y	Y	Y	Y
Memory scrubbing	Y	Y	Y	Y	Y	Y
Chipkill memory	Y	Y	Y	Y	Y	Y
Memory page deallocation	Y	Y	Y	Y	N	N
L1 parity check plus retry	Y	Y	Y	Y	Y	Y
L2 cache line delete	Y	Y	Y	Y	Y	Y
Array recovery & Array persistent deallocation (spare bits in L1 & L2 cache; L1, L2 directory)	Y	Y	Y	Y	Y	Y
Special uncorrectable error handling	Y	Y	Y	Y	Y	Y
Fault Detection and Isolation						
Platform FFDC diagnostics	Y	Y	Y	Y	Y	Y
I/O FFDC diagnostics	Y	Y	Y	Y	N	Y
Runtime diagnostics	Y	Y	Y	Y	Limited	Limited
Storage protection keys	Y	Y	Y	Y	N	N

RAS feature	AIX V5.3	AIX V6.1	IBM i 5.4 LIC 5.4.5	IBM i 6.1	RHEL V5.1	SLES V10
Dynamic trace	N	Y	Y	Y	N	N
Operating system FFDC	Y	Y	Y	Y	N	N
Error log analysis	Y	Y	Y	Y	Y	Y
Service processor support for BIST for logic and arrays, wire tests, and component initialization	Y	Y	Y	Y	Y	Y
Serviceability						
Boot time progress indicator	Y	Y	Y	Y	Limited	Limited
Firmware error codes	Y	Y	Y	Y	Y	Y
Operating system error codes	Y	Y	Y	Y	Limited	Limited
Inventory collection	Y	Y	Y	Y	Y	Y
Environmental and power warnings	Y	Y	Y	Y	Y	Y
Hot plug fans, power supplies	Y	Y	Y	Y	Y	Y
Extended error data collection	Y	Y	Y	Y	Y	Y
SP call home on non-HMC configurations	Y	Y	Y	Y	Y	Y
I/O drawer redundant connections	Y	Y	Y	Y	Y	Y
I/O drawer hot-add and concurrent repair	Y	Y	Y	Y	Y	Y
SP mutual surveillance with POWER Hypervisor	Y	Y	Y	Y	Y	Y
Dynamic firmware update with the HMC	Y	Y	Y	Y	Y	Y
Service agent call home application	Y	Y	Y	Y	Y	Y
Guiding light LEDs	Y	Y	Y	Y	Y	Y
System dump for memory, POWER Hypervisor, SP	Y	Y	Y	Y	Y	Y
Operating system error reporting to HMC SFP application	Y	Y	Y	Y	Y	Y
RMC secure error transmission subsystem	Y	Y	Y	Y	Y	Y
Health check scheduled operations with HMC	Y	Y	Y	Y	Y	Y
Operator panel (virtual or real)	Y	Y	Y	Y	Y	Y
Redundant HMCs	Y	Y	Y	Y	Y	Y
Automated recovery/restart	Y	Y	Y	Y	Y	Y
Repair and verify guided maintenance	Y	Y	Y	Y	Limited	Limited
Concurrent kernel update	N	Y	Y	Y	N	N

1. Feature is not supported on Version 4 of RHEL

4.5 Manageability

Several functions and tools help manageability, and can allow you to efficiently and effectively manage your system.

4.5.1 Service processor

The service processor is a controller running its own operating system. It is a component of the service interface card.

The service processor operating system has specific programs and device drivers for the service processor hardware. The host interface is a processor support interface connected to the POWER6 processor. The service processor is always working, regardless of main system unit's state. The system unit can be in the following states:

► Standby (Power off)
► Operating, ready to start partitions
► Operating with running logical partitions

The service processor is used to monitor and manage the system hardware resources and devices. The service processor checks the system for errors, ensuring the connection to the HMC for manageability purposes and accepting Advanced System Management Interface (ASMI) Secure Sockets Layer (SSL) network connections. The service processor provides the ability to view and manage the machine-wide settings using the ASMI, and enables complete system and partition management from the HMC.

> **Note:** The service processor enables a system that will not boot to be analyzed. The error log analysis can be performed from either the ASMI or the HMC.

The service processor uses 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 ASMI. The ASMI options can be accessed through an HTTP server that is integrated into the service processor operating environment.
► Both Ethernet ports have a default IP address:
  – Service processor Eth0 or HMC1 port is configured as 169.254.2.147
  – Service processor Eth1 or HMC2 port is configured as 169.254.3.147

4.5.2 System diagnostics

The 8203-E4A (and 8204-E8A) provide system diagnostics including stand-alone diagnostics, which are loaded from the DVD-ROM drive, and online diagnostics (available in AIX).

► Online diagnostics, when installed, are a part of the AIX operating system 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 error log and the AIX configuration data.
  – Service mode, which requires a service mode boot of the system, enables the checking of system devices and features. Service mode provides the most complete checkout of the system resources. All system resources, except the SCSI adapter and the disk drives used for paging, can be tested.

- – Concurrent mode enables the normal system functions to continue while selected resources are being checked. Because the system is running in normal operation, some devices might require additional actions by the user or diagnostic application before testing can be done.

- – Maintenance mode enables the checking of most system resources. Maintenance mode provides the 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 on the SMS menus. This error log contains errors that are found by partition firmware when the system or partition is booting.

► The service processor's error log can be accessed on the ASMI menus.

► You can also access the system diagnostics from a Network Installation Management (NIM) server.

> **Note:** When a Power 520 system is ordered with AIX or Linux as the primary operating system, a DVD-ROM (#5756) or DVD-RAM (#5757) is optional. An alternate method for maintaining and servicing the system must be available if you do not order the DVD-ROM or DVD-RAM

The 9407-M15, 9408-M25 (and 9409-M50) provide Dedicated Service Tools (DST) as part of the IBM i licensed machine code (Licensed Internal Code) and System Service Tools (SST) as part of the IBM i operating system. DST can be run in dedicated mode (no operating system loaded). DST tools and diagnostics are a superset of those available under SST.

The IBM i End Subsystem (ENDSBS *ALL) command can shut down all IBM and customer applications subsystems except the controlling subsystem QTCL. The Power Down System (PWRDWNSYS) command can be set to power down the IBM i partition and restart the partition in DST mode.

You can start SST during normal operations, which leaves all applications up and running, using the IBM i Start Service Tools (STRSST) command (when signed onto IBM i with the appropriately secured user ID).

With DST and SST you can look at various logs, run various diagnostics, or take several kinds of system dumps, or take other options. The service level functions you would typically see when using the service tools interface are:

► Product activity log
► Trace Licensed Internal Code
► Work with communications trace
► Display/Alter/Dump
► Licensed Internal Code log
► Main storage dump manager
► Hardware service manager

4.5.3 Electronic Service Agent

Electronic Service Agent™ and the IBM Electronic Services Web portal comprise the IBM Electronic Service solution. IBM Electronic Service Agent is a no-charge tool that proactively monitors and reports hardware events, such as system errors, performance issues, and inventory. Electronic Service Agent can free the clients to focus on their company strategic

business initiatives, save time, and spend less effort managing day-to-day IT maintenance issues.

The no-additional-charge Electronic Service Agent software is available on IBM operating systems and some non-IBM operating systems. It has to be set up and activated on each supported system or partition within a system. An associated IBM service ID can be established to non-IBMers with proper authorization to review service-based information collected from IBM systems. The agent:

► Provides 24 x 7 System Monitoring
► Reports hardware error log entries and performance information
► Automates hardware problem submission
► Tracks system hardware and software inventory
► Automates Microcode PTF downloads
► Provides the IBM Customer Service Representative (CSR) with access to data while diagnosing problems

Secured information retrieval and secured access to the collected information is provided. An IBM service representative and properly registered customer personnel can review and collaborate over the information stored at the IBM Electronic Services Web. This Web site offers:

► A single entry point for hardware and software support
► 24-hour access to your customized Information Technology (IT) information
► Access to Web-delivered Premium Services
  (Premium search is a service using information gathered by Service Agents. This is an additional charge service that requires a special contract.)
► The ability to submit a hardware and/or software problem electronically
► The ability to research technical problems
► The ability to view Electronic Service Agent collected information

The primary focus areas of Electronic Service Agent are to provide an automatic problem-reporting function and enable review of collected hardware and software inventory. The problem reporting helps predict and prevent hardware errors by early detection of potential problems. Assuming you have the communication link to IBM set up, a problem can be reported to IBM and, if appropriate, a replacement part sent or an IBM service representative dispatched to your location.

IBM Electronic Service Agent can also report software errors, initiate the download of fixes, and automatically submit a problem to IBM when appropriate. The categories of collected information include more than just problem-related information. The range of collected information includes:

► System resource utilization
► Performance
► System failure logs
► Feature codes
► Part numbers, serial numbers
► Part locations
► Software inventory
► Operating system applications
► Program temporary fix (PTF)
► Maintenance levels
► Configuration values

Supplemental system service information is sent to IBM and made available to IBM support centers to aid in problem resolution. The same service information is available to be viewed

with prior registration of an IBM ID. For more information about creating an IBM ID, registering your system, and viewing the service information that you share with IBM, go to the Web site.

The Electronic Service Agent is offered for IBM z and IBM x systems as well as System p, System i, and Power systems servers. It is integrated into AIX starting with AIX V5.3 TL6 (and later) and has been integrated with IBM i for many releases and continues on 5.4 and 6.1.

IBM also provides Electronic Service Agent for the following Linux and Windows operating systems:

► Linux Red Hat Enterprise Advanced Server 4
► Linux SUSE 9.0
► Microsoft Windows XP with service pack (SP) 1
► Microsoft Windows 2000 Server
► Microsoft Windows 2003 Server

For some systems, you can have multiple logical partitions with one or more operating systems running. No matter which system you have, how many logical partitions are configured, or which operating systems are running, IBM Electronic Service Agent should be installed (no charge, but can be selectively un-installed) and activated on each logical partition, each operating system, and each HMC.

On an AIX or Linux partition Electronic Service Agent V6 can be activated by typing `smitty esa_main` and selecting "Configure Electronic Service Agent." Additionally, ESA V6 includes a powerful Web-based user interface, giving the administrator easy access to status, tool settings, and problem information.

You can manage and control Electronic Service Agent and its functions through two user interfaces under IBM i. Service Agent problem reporting must be managed through the IBM i command line (enter GO SERVICE). Windows operating system based System i Navigator can be used to collect and transmit Electronic Service Agent service information for a group of systems as set up under the Management Central interfaces.

The IBM i Information Center for 5.4 and 6.1 has much documentation on setting up and using Electronic Customer Support and Electronic Service Agent.

You set up, manage, and control what Electronic Service Agent functions are used, depending upon the operating system interface of the agent and what you authorize.

The functions provided by IBM Electronic Service Agent vary based on the network environment. For example, if your system is running two AIX logical partitions and two IBM i logical partitions, the HMC reports platform errors for the system and logical partition errors for the AIX logical partitions. IBM Electronic Service Agent on each IBM i logical partition reports partition errors for the logical partition on which it is running. For this reason, to have total coverage (assuming only IBM i and AIX partitions are active), IBM Electronic Service Agent needs to be running on each HMC, each AIX logical partition, and each IBM i logical partition.

IBM Electronic Services has functions and interfaces to ensure privacy and security between IBM clients and IBM for electronic communications. System inventory information is treated as confidential information and stored in a secure database that is protected behind IBM firewalls.

Electronic Service Agent securely transmits via the Internet (HTTPS or VPN) and can be configured to communicate securely through gateways to provide clients a single point of exit from their site. Communication between the client and IBM only flows one way. Activating

Service Agent does not enable IBM to call into a client's system. The client's business applications or business data is never transmitted to IBM

For more information about ESA security, see the HMC Security Whitepaper available at:

> `http://www14.software.ibm.com/webapp/set2/sas/f/best/home.htm`]

There are many additional sources of information to help you set up and use IBM Electronic Service Agent. Some of them include:

AIX Information Center:

> `http://publib.boulder.ibm.com/infocenter/systems/scope/aix/index.jspsao`

> Select your release level, for example, AIX 6, and search for "electronic service agent."

IBM i (i5/OS) Information Center:

> `http://publib.boulder.ibm.com/infocenter/systems/scope/i5os/index.jsp`

> **Note:** To receive maximum coverage, activate Electronic Service Agent on every platform, partition, and Hardware Management Console (HMC) in your network. If your IBM server is managed by an HMC, the HMC can be set up to report all hardware problems, and the operating system will report only software problems and system information. You must configure the Electronic Service Agent on the HMC. The AIX operating system will not report hardware problems for a system managed by an HMC.

IBM Electronic Service provide these benefits:

Increased uptime   Electronic Service Agent is designed to enhance warranty and maintenance service by providing faster hardware error reporting and uploading system information to IBM support. 24x7 monitoring and reporting means no more dependency on human intervention or off-hours client personnel when errors are encountered in the middle of the night.

Security   Electronic Service Agent is secure in monitoring, reporting, and storing the data at IBM. Electronic Service Agent securely transmits via the Internet (HTTPS or VPN) and can be configured to communicate securely through gateways to provide clients a single point of exit from their site. Communication between the client and IBM only flows one way. Activating Service Agent does not enable IBM to call in to a client's system. System inventory information is stored in a secure database, which is protected behind IBM firewalls. The client's business applications or business data is never transmitted to IBM

More accurate reporting

      Since system information and error logs are automatically uploaded to the IBM support Center in conjunction with the service request, clients are not required to find and send system information, decreasing the risk of misreported or misdiagnosed errors. Once inside IBM, problem error data is run through a data knowledge management system and knowledge articles are appended to the problem record.

Customized support   Using the IBM ID entered during activation, clients can view system and support information in the "My Systems" and "Premium Search" sections of the Electronic Services Web site.

The Electronic Services Web portal is a single Internet entry point that replaces the multiple entry points traditionally used to access IBM Internet services and support. This Web portal

enables you to gain easier access to IBM resources for assistance in resolving technical problems.

Service Agent provides these additional services:

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

https://www-304.ibm.com/jct03004c/support/electronic/portal

► Premium Search: A search service using information gathered by Service Agents (this is a paid service that requires a special contract).

For more information on how to utilize the power of IBM Electronic Services, visit the following Web site or contact an IBM Systems Services Representative:

https://www-304.ibm.com/jct03004c/support/electronic/portal

4.5.4 Manage serviceable events with the HMC

Service strategies become more complicated in a partitioned environment. The "Manage Serviceable Events" task in the HMC can help streamline this process.

Each logical partition reports errors it detects, without determining whether 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.

By using the Manage Serviceable Events task in the HMC, you can avoid long lists of repetitive call-home information by recognizing that these are repeated errors and consolidating them into one error.

In addition, you can use the Manage Serviceable Events task to initiate service functions on systems and logical partitions, including the exchanging of parts, configuring connectivity, and managing dumps.

4.5.5 Hardware user interfaces

Advanced System Management Interface

The Advanced System Management interface (ASMI) is the interface to the service processor that enables you to manage 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. Some repair procedures require connection to the ASMI.

The ASMI is accessible through the HMC. It is also accessible using a Web browser on a system that is connected directly to the service processor (in this case, either a standard Ethernet cable or a crossed cable) or through an Ethernet network. Use the ASMI to change the service processor IP addresses or to apply some security policies and prevent access from undesired IP addresses or ranges.

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 an HMC

If configured to do so, the HMC connects directly to the ASMI for a selected system from this task.

To connect to the Advanced System Management interface from an HMC:

1. Open Systems Management from the navigation pane.

2. From the work pane, select one or more managed systems to work with.

3. From the System Management tasks list, select Operations → Advanced System Management (ASM).

Accessing the ASMI using a Web browser

The Web interface to the ASMI is accessible through Microsoft Internet Explorer 6.0, Microsoft Internet Explorer 7, 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 (IPL) and run time. However, some of the menu options in the Web interface are unavailable during IPL or run time to prevent usage or ownership conflicts if the system resources are in use during that phase. The ASMI provides a Secure Sockets Layer (SSL) Web connection to the service processor. To establish an SSL connection, open your browser using `https://`.

> **Note:** To make the connection through Internet Explorer, click Tools → Internet Options. Uncheck "Use TLS 1.0", and click OK.

Accessing the ASMI using an ASCII terminal

The ASMI on an ASCII terminal 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 IPL and run time.

Graphics terminal

The graphics terminal is available to users who want a graphical user interface (GUI) to their AIX or Linux systems. To use the graphics terminal, plug the graphics adapter into a PCI slot in the back of the server. You can connect a standard monitor, keyboard, and mouse to the adapter to use the terminal. This connection allows you to access the SMS menus, as well as an operating system console.

4.5.6 IBM Power Systems firmware maintenance

The IBM Power Systems Client-Managed Microcode is a methodology that enables you to manage and install microcode updates on Power Systems, IBM System p, IBM System p5, pSeries, and RS/6000 systems and associated I/O adapters. The IBM POWER5 or POWER6 server firmware and microcode can be downloaded and installed either from an HMC or from a running partition in case that system is not managed by an HMC.

For access to the initial Web pages addressing this capability see:

http://www.ibm.com/systems/support

For POWER6 technology systems, select the **Power** link. Figure 4-7 on page 189 is an example of the next window displayed.
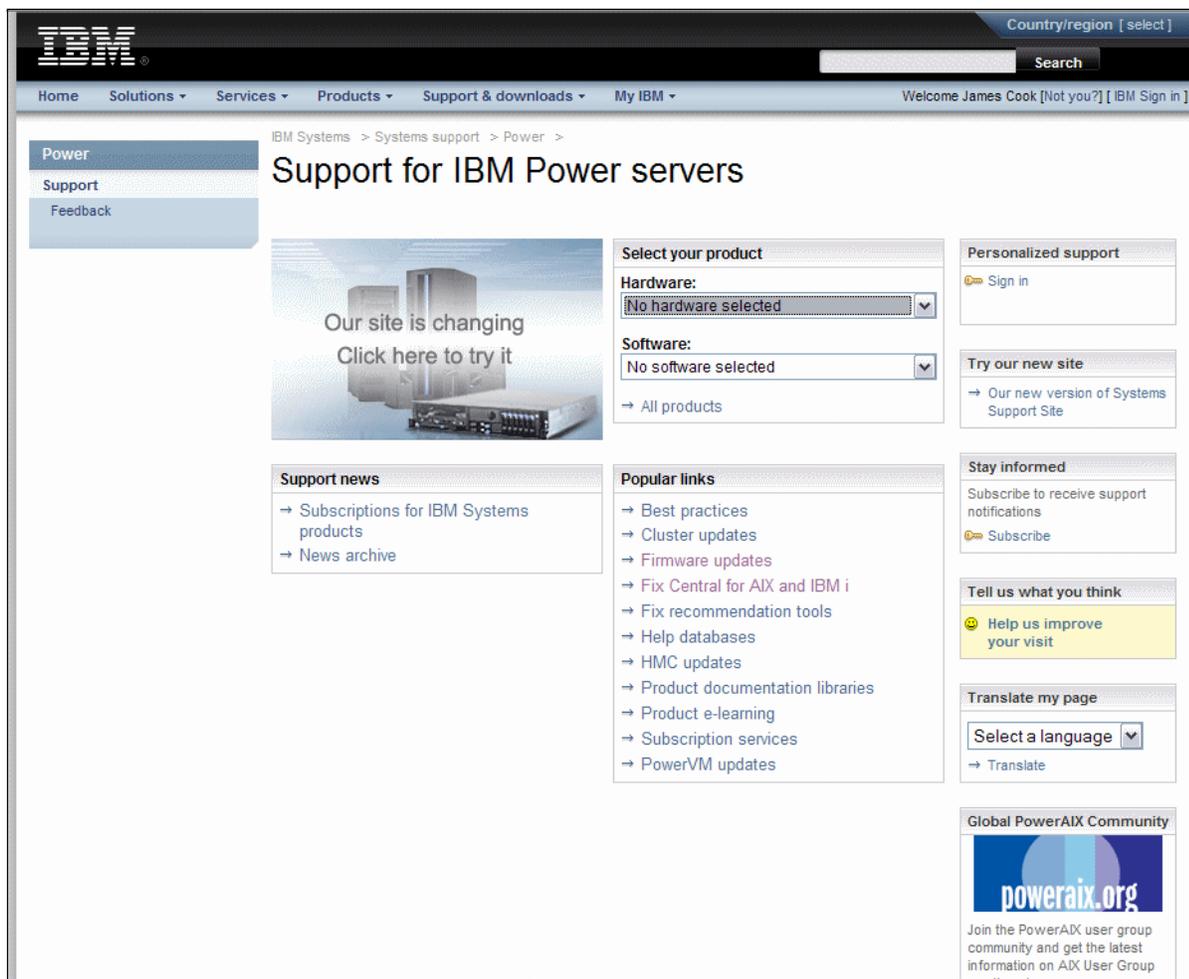
Figure 4-7   Support for IBM Power servers window

► Under the Hardware category, select the appropriate MTM and follow the links to access a variety of downloads.

We selected 9408-M25 MTM; then selected the Download tab, from which we could further select:

– Firmware updates

– Fix Level Recommendations tool (AIX)

– HMC update

– Recommended firmware fixes (IBM i)

► Under the Software category, select the appropriate operating system or HMC.

► Under the Popular links heading, we select Firmware updates. The next window gave us more selections, one of which was system firmware for 82nn and 91nn MTMs.

Note, firmware level EL340_039 or later is required to support hardware-based announcements made during 4Q 2008. This includes 8203_E4A support of the IBM i operating system.

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

Each IBM POWER5 or POWER6 server has the following levels of server firmware and power subsystem firmware:

► Installed level – This is the level of server firmware or power subsystem firmware that has been installed and will be installed into memory after the managed system is powered off and powered on. It is installed on the t side of system firmware.

► Activated level – This is the level of server firmware or power subsystem firmware that is active and running in memory.

► Accepted level – This is the backup level of server or power subsystem firmware. You can return to this level of server or power subsystem firmware if you decide to remove the installed level. It is installed on the p side of system firmware.

IBM provides the Concurrent Firmware Maintenance (CFM) function on System p systems. This function supports applying nondisruptive system firmware service packs to the system concurrently (without requiring a reboot to activate changes). For systems that are not managed by an HMC, the installation of system firmware is always disruptive.

The concurrent levels of system firmware can, on occasion, contain fixes that are known as *deferred*. These deferred fixes can be installed concurrently but are not activated until the next IPL. For deferred fixes within a service pack, only the fixes in the service pack, which cannot be concurrently activated, are deferred. Figure 4-8 shows the system firmware file naming convention.


• PPNNSSS_FFF_DDD

ASMI Firmware Level Display

PP	Package Identifier	01	Managed System
		02	Power Subsystem
NN	Platform & Class	EL	Low End
		EM	Mid Range
		EH	High End
		EB	Bulk Power
SSS	Release Indicator		
FFF	Current Fixpack		
DDD	Last Distruptive Fixpack		

EM310_023

Figure 4-8   Firmware file naming convention

Here is one example:

```
01EM310_026_026 = Managed System Firmware for 9117-MMA Release 310 Fixpack 026
```

An installation is disruptive if:

► The release levels (*SSS*) of currently installed and new firmware are different.

► The service pack level (*FFF*) and the last disruptive service pack level (*DDD*) are equal in new firmware.

Otherwise, an installation is concurrent if:

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

4.5.7 Management Edition for AIX

IBM Management Edition for AIX (ME for AIX) is designed to provide robust monitoring and quick time to value by incorporating out-of-the box best practice solutions that were created by AIX and PowerVM Virtual I/O Server developers. These best practice solutions include predefined thresholds for alerting on key metrics, Expert Advice that provides an explanation of the alert and recommends potential actions to take to resolve the issue, and the ability to take resolution actions directly from the Tivoli Enterprise Portal or set up automated actions. Users have the ability to visualize the monitoring data in the Tivoli Enterprise Portal to determine the current state of the AIX, LPAR, CEC, HMC, and VIOS resources.

Management Edition for AIX is an integrated systems management offering created specifically for the system p platform that provides, as primary functions, the following capabilities:

► Monitoring of the health and availability of the system p

► Discovery of configurations and relationships between system p service and application components

► Usage and accounting of system p IT resources

For information regarding the ME for AIX, visit this link:

> http://www-03.ibm.com/systems/p/os/aix/sysmgmt/me/index.html

4.5.8 IBM Director

IBM Director versions and related components have been available for several years, providing a wide range of hardware and software event monitoring, performance monitoring, software maintenance, and more functions across AIX, IBM i, Linux, and other operating systems.

IBM Systems Director 6.1 is the latest version. It provides a platform management foundation that streamlines the way you manage physical and virtual systems across a heterogeneous environment. By using industry standards, IBM Systems Director supports multiple operating systems and virtualization technologies across IBM and non-IBM x86 platforms.

Through a single user interface, IBM Systems Director provides consistent views for viewing managed systems, determining how these systems relate to one another, and identifying their statuses, thus helping to correlate technical resources with business needs. A set of common tasks included with IBM Systems Director provides many of the core capabilities required for basic management, which means instant out-of-the-box business value. These common tasks include discovery, inventory, configuration, system health, monitoring, updates, event notification, and automation across managed systems.

IBM Systems Director's Web and command-line interfaces provide a consistent interface focused on driving these common tasks and capabilities:

► Discovering, navigating, and visualizing systems on the network with the detailed inventory and relationships to the other network resources

► Notifying users of problems that occur on the system and enabling the ability to drill down to the source of the problem

► Notifying users when systems need updates and distributing and installing updates on a schedule

► Analyzing real-time data for systems and setting critical thresholds that notify the administrator of emerging problems

- ► Configuring settings of a single system and creating a configuration plan that can apply those settings to multiple systems

- ► Updating installed plug-ins to add new features and function to the base capabilities

- ► Managing the life cycle of virtual resources

- ► Can be extended to provide access to additional tools from the same user interface

For information regarding the IBM Director, visit the following link:

http://www-03.ibm.com/systems/management/director/

4.6 Cluster-based availability solutions

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:

- ► High processing capacity

- ► Resource consolidation

- ► Optimal use of resources

- ► Geographic server consolidation

- ► 24x7 availability with failover protection

- ► Disaster recovery

- ► Scale-out and scale-up without downtime

- ► Centralized system management

IBM augments the IBM hardware availability capabilities with a range of software availability functions under the supported operating systems and expanded availability functions available with IBM products built on top of the operating system support.

This topic focuses on the primary IBM higher availability software products that are built on top of the operating systems.

The primary IBM availability Web site for Power Systems contains more detailed information than is summarized within this section. You can access it at:

http://www.ibm.com/systems/power/software/availability/

Note: There are additional non-IBM higher availability products that run on these operating systems. Discussion of these non-IBM products is beyond the scope of this publication. You can link to some of these non-IBM products from the main Power System availability Web site.

4.6.1 AIX and Linux availability solutions

This section discusses two major categories of IBM products for providing availability and management of multiple systems running AIX and Linux:

- ► IBM PowerHA suite of products and offerings

- ► IBM Cluster Systems Management software (CSM) offerings

IBM PowerHA suite of products and offerings

For several years IBM High Availability Clustered Multi Processing (HACMP) has provided reliable high availability services, monitoring clients' mission critical applications running on IBM servers—and now the IBM System p and System i servers. From the network server operating systems and applications, HACMP monitors the entire system and can automatically restart an application on backup hardware in the event of failure or service degradation. The combination of IBM's servers, AIX or Linux for Power, and HACMP provides the highest level of protection and availability.

PowerHA is the new name for HACMP for AIX and Linux offerings.

PowerHA for AIX (HACMP)

PowerHA for AIX, formerly IBM High Availability Cluster Multi-Processing (HACMP) is now at level V5.4.1. This release level offers robust high availability and disaster recovery for IBM System p and System i customers with mission-critical applications.

New HACMP V5.4.1 features include:

► AIX Workload Partitions (WPAR).

► HACMP/XD support for IBM TotalStorage disk subsystem (PPRC), including Consistency Groups.

► New GLVM monitoring. A new Geographical Logical Volume Manager (GLVM) Status Monitor provides the ability to monitor GLVM status and state. These monitors enable you to keep better track of the status of your application data when using the HACMP/XD GLVM option for data replication.

► Improved support for NFS V4, which includes additional configuration options, as well as improved recovery time. HACMP can support both NFS V4 and V2/V3 within the same high availability environment.

► NFSv4 support improvements.

► HACMP usability and RAS improvements.

► New options for detecting and responding to a partitioned cluster.

The optional features HACMP/XD and HACMP Smart Assist for AIX V6.1 provide high availability disaster recovery solutions for your business.

HACMP Smart Assists for DB2 and Oracle simplify HACMP setup in these common database environments. Including the Smart Assist for WebSphere, these Smart Assists use application-specific knowledge to extend HACMP's standard auto-discovery features and provide all the necessary application monitors and start/stop scripts to streamline the configuration process. All three Smart Assists are included in one inexpensive, orderable package.

PowerHA XD for AIX (HACMP/XD)

PowerHA XD for AIX, formerly HACMP/XD (Extended Distance), extends HACMP's capabilities by replicating critical data and enabling failover to a remote site. HACMP/XD provides a portfolio of data mirroring and recovery options which let you build a disaster recovery solution with the backup site in a campus- or metropolitan-wide area, or even hundreds of miles away.

HACMP's Extended Distance option (HACMP/XD) extends the protection of HACMP for AIX to geographically remote sites to help ensure business continuity even if an entire site is disabled by catastrophe. HACMP/XD automatically manages the replication and synchronization of your Web site, databases, and other critical data and applications to a separate location from your primary operations and keeps this replicated data updated in real

time. In the event of a power outage, flood, or other disaster, your services are automatically transferred from the disabled site and restarted at the remote site; your data remains safe and your business keeps running. During operation at the backup site, data replicating will be reversed back to the primary site as soon as it is restored, providing protection for the backup site and enabling the primary site to resume production operations if desired.

PowerHA for Linux (HACMP for Linux)

IBM's HACMP premiere high availability product is also available for use on the Linux operating system. HACMP V5.4 for Linux supports Red Hat and SUSE Linux for Power systems. With HACMP for Linux, clients can have a common availability solution throughout their enterprise, from high-end servers running a large database to edge-tier applications on Linux. This extends HACMP's world-class protection to your mission-critical applications running on all IBM System p servers. HACMP for Linux uses the same interface and configurations as HACMP for AIX, providing a common multi-platform solution which protects your investment as you grow your cluster environments.

Cluster Systems Management products and offerings

These software and offerings are targeted toward POWER processor-based AIX and Linux cluster target scientific and technical computing, large-scale databases, and workload consolidation environments. They provide a comprehensive tool designed specifically to streamline initial deployment and ongoing management of cluster systems.

IBM Cluster Systems Management software (CSM) is designed to help reduce the overall cost and complexity of IT management by simplifying the tasks of installing, configuring, operating, and maintaining clusters of servers or logical partitions (LPARs). CSM offers a single consistent interface for managing both AIX and Linux nodes, with capabilities for remote parallel network installation, remote hardware control, distributed command execution, file collection and distribution, cluster-wide monitoring capabilities, and integration with High Performance Computing (HPC) applications.

CSM V1.7, which is needed to support POWER6 processor-based HMC, includes Highly Available Management Server (HA MS®) at no additional charge. CSM HA MS is positioned for enterprises that need a highly available management server. CSM HA MS is designed to remove the management server as a single point of failure in the cluster.

Two important IBM CSM products are IBM System Cluster 1600 and IBM System Cluster 1350™.

IBM System Cluster 1600

IBM System Cluster 1600 solutions are comprised of IBM System p based servers running either AIX or Linux. Cluster 1600 is a highly scalable cluster solution intended for use in configurations supporting large-scale computational modeling and analysis, large databases and business intelligence applications, and cost-effective datacenter, server, and workload consolidation.

Cluster 1600 systems can be deployed on Ethernet networks, InfiniBand networks, or with the IBM High Performance Switch and are typically managed with Cluster Systems Management (CSM) software.

Some key characteristics of IBM System Cluster 1600 include:

► Cluster Systems Management (CSM) software for comprehensive, flexible deployment and ongoing management

► Cluster interconnect options: industry standard 1/10Gb Ethernet (AIX or Linux), IBM High Performance Switch (AIX), 4x/12x InfiniBand (AIX or Linux), or Myrinet (Linux)

- ► Operating system options: AIX Version 5.2/5.3/6.1, SUSE Linux Enterprise Server 9 or 10, Red Hat Enterprise Linux 4 or 5

- ► Complete software suite for creating, tuning, and running parallel applications: Engineering & Scientific Subroutine Library (ESSL), Parallel ESSL, Parallel Environment, XL Fortran, VisualAge® C++

- ► High-performance, high availability, highly scalable cluster file system General Parallel File System (GPFS™)

- ► Job scheduling software to optimize resource utilization and throughput: LoadLeveler®

- ► High availability software for continuous access to data and applications: IBM PowerHA for AIX and IBM PowerHA for Linux

IBM System Cluster 1350

IBM high-performance computing (HPC) clustered solutions offer significant price/performance advantages for many high-performance workloads by harnessing the advantages of low-cost servers plus innovative, easily available, open source and commercially available software – all running on hardware other than the Power 520 and 550.

IBM has designed the IBM System Cluster 1350 to help address installation and management tasks of this environment. Using advanced Intel Xeon®, AMD™ Opteron™, and IBM POWER6-based Blade server nodes, proven cluster management software, and optional high-speed interconnects, the Cluster 1350 offers the best of IBM and third-party technologies. As a result, clients can speed up installation of an HPC cluster, simplify its management and support, and reduce mean time to payback.

IBM System Cluster 1350 characteristics include:

- ► Rack-optimized Intel Xeon dual-core and quad-core and AMD Opteron processor-based servers

- ► Intel Xeon, AMD Opteron and POWER6-based blades

- ► Optional high capacity IBM System Storage DS3200, DS3400, DS4700, DS4800, and EXP3000 Storage Servers, and IBM System Storage EXP 810 Storage Expansion

- ► Optional high performance Myricom, BNT™, or Force 10 10Gb cluster interconnect

- ► Optional Cisco, Voltaire, and QLogic® DDR InfiniBand cluster interconnects

- ► Runs with RHEL5 or SLES10 Linux operating systems or Windows Compute Cluster Server 2003 (Windows HPC Server 2008 in August 2008)

- ► Scales up to 1,024 cluster nodes

- ► Optional Linux or Windows cluster installation and support services from IBM Global Services or an authorized partner or distributor

For more information on IBM System Cluster 1350 solutions, see:

http://www-03.ibm.com/systems/clusters/hardware/1350/

4.6.2 IBM i availability solutions

IBM i in releases 5.4 and especially 6.1 itself provides many ways to back up important system and applications components and data. IBM i 6.1 provides further extensions to the "save while active" status of objects. High availability business partners have provided a rich set of high availability products and offerings in this area.

During January 2008 IBM introduced the System i High Availability Solutions Manager (HASM). This product builds on IBM i 6.1 operation system's set of "base availability enhancements." During 2008 HASM was renamed as IBM PowerHA for i.

PowerHA for i

PowerHA for i is the IBM high availability disk-based clustering solution for the IBM i 6.1 operating system. PowerHA for i when combined with IBM i clustering technology delivers a complete high availability and disaster recovery solution for your business applications running in the IBM System i environment. PowerHA for i enables you to support high availability capabilities with either native disk storage or the IBM DS8000 or DS6000 storage servers.

PowerHA for i provides a complete high-availability solution based on:

► IBM i's integrated support for independent auxiliary storage pools (IASPs)
► Using IBM i cross-site mirroring (XSM)
► IBM Storage Solutions Metro Mirror, Global Mirror, FlashCopy®
► IBM i system administrative domain capabilities.

Both graphical browser-based and IBM i command level interfaces are supported:

► GUI to manage clusters, resource groups, and domains
► 5250 command interface for command-line management

Requirements include:

► Systems must be running IBM i 6.1/OS V6R1
► Applications and data must be deployed using IASPs

IBMHA for i also includes integrated source- and target-side tracking for XSM and geographic mirroring. This means that when you detach a target system, the resynchronize operation, after reattaching, includes only the changed objects on the source and target system.

XSM enables you to perform role-swap operations using Metro Mirror, a synchronous replication product for the DS8000 server. You can readily perform both planned and unplanned switching operations with minimal impact on operational procedures. You should use Metro Mirror for best-case recovery point objective (RPO) and recovery time objective (RTO).

To learn more about this product refer to *Implementing PowerHA for IBM i*, SG24-7405.

IBM iCluster for i

iCluster is the IBM i HA/DR solution based on logical replication. iCluster is based upon IBM i journaling technology and is optimized for near real time replication and geographic dispersion. iCluster provides flexibility for those customers looking to use the second copy of the data on the backup system for read-only access. iCluster's MatchMerge technology and advanced sync check capabilities are designed to enhance data consistency.

The iCluster portfolio consists of the following products:

► iCluster: This is a logical replication solution based on IBM i journaling, designed for large enterprises with more complex data availability and recoverability requirements.

► iBalance: This is an additional functions chargeable feature of iCluster. The key additional functions are focused on bidirectional replication for workload balancing.

► iCluster: This has a basic range of functions intended for small and medium business (SMB) organizations with a simpler set of requirements.

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 publications

For information about ordering these publications, see "How to get Redbooks publications" on page 199. Note that some of the documents referenced here may be available in softcopy only.

► *PowerVM Virtualization on IBM System p: Introduction and Configuration Fourth Edition*, SG24-7940

► *PowerVM Virtualization on IBM System p: Managing and Monitoring*, SG24-7590

► *Getting Started with PowerVM Lx86*, REDP-4298

► *PowerVM Live Partition Mobility on IBM System p*, SG24-7460

► *Integrated Virtualization Manager on IBM System p5*, REDP-4061

► *Introduction to Workload Partition Management in IBM AIX Version 6.1*, SG24-7431

► *Hardware Management Console V7 Handbook*, SG24-7491

► *IBM System i and System p*, SG24-7487

► *IBM System p 550 Technical Overview and Introduction*, REDP-4404

► *IBM System p 570 Technical Overview and Introduction*, REDP-4405

► *IBM System i Overview: Models 515, 525, 550, 570, 595, and More*, REDP-5052

► *IBM Power 520 and Power 550 (POWER6) System Builder*, REDP-4412

► *Case Study: Architecting SOA Solutions for Changing Economic Environments*, REDP-4414

► *IBM BladeCenter JS12 and JS22 Implementation Guide*, SG24-7655

Other publications

These publications are also relevant as further information sources for planning:

► *Logical Partitioning Guide*, SA76-0098

► *Site and Hardware Planning Guide*, SA76-0091

► *Site Preparation and Physical Planning Guide*, SA76-0103

These publications are also relevant as further information sources for installing:

► *Installation and Configuration Guide for the HMC*, SA76-0084

► *Power Systems PCI Adapter Placement Guide for Machine Type 820x and 91xx*, SA76-0090.

► *Power Systems PCI Adapter Placement Guide for Machine Type 940x,* SA76-0096

These publications are also relevant as further information sources for using your system:

- ► *Introduction to Virtualization*, SA76-0145
- ► *Operations Guide for the ASMI and for Nonpartitioned Systems*, SA76-0094
- ► *Operations Guide for the HMC and Managed Systems*, SA76-0085
- ► *Virtual I/O Server Command Reference*, SA76-0101

These publications are also relevant as further information sources for troubleshooting:

- ► *AIX Diagnostics and Service Aids*, SA76-0106
- ► *Managing Devices*, SA76-0107
- ► *Managing PCI Devices*, SA76-0092
- ► *SAS RAID Controller Reference Guide*, SA76-0112
- ► *Service Guide for HMC Models 7042-Cr4 and 7042-C06*, SA76-0120
- ► Performance information, including the *IBM Power Systems Performance Capabilities Reference IBM i operating system Version 6.1*, which is available at:

  http://www-03.ibm.com/systems/resources/pcrm.pdf

  or

  http://www-03.ibm.com/systems/i/advantages/perfmgmt/resource.html
- ► *System i and System p: Advanced POWER Virtualization Operations Guide*, SA76-0100-03. or later. This publication contains detailed information, including sizing for performance the capabilities of functions such as the Virtual I/O Server (VIOS) partitions and communications traffic handled by the Integrated Virtual Ethernet adapter (IVE). The IVE is also known as the Host Ethernet Adapter (HEA).

  You can find this publication at the IBM Systems Hardware Information Center at:

  http://publib.boulder.ibm.com/infocenter/systems/scope/hw

  Search with Advanced POWER Virtualization Operations Guide

  Select the topmost PowerVM Editions link

  Scroll down to find "For more information about PowerVM Editions, see the PowerVM Editions Operations Guide." Click the link to the PDF.

Online resources

These Web sites are also relevant as further information sources:

- ► IBM Systems Information Center

  This is a very good starting place for information centers on IBM hardware, systems software (such as IBM Director), and the supported operating systems.

  http://publib.boulder.ibm.com/infocenter/systems
- ► Support for IBM System p

  http://www.ibm.com/systems/support/p
- ► Support for IBM System i

  http://www.ibm.com/systems/support/i
- ► IBM System Planning Tool

  http://www.ibm.com/systems/support/tools/systemplanningtool

- Fix Central for IBM operating systems, IBM hardware platforms, and IBM software maintenance packages downloads

  http://www.ibm.com/eserver/support/fixes

- Microcode downloads

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

- Linux for IBM Power Systems, System p and System i

  http://www-03.ibm.com/systems/power/software/linux/

IBM System Storage Interoperation Center (SSIC) Web site

  http://www-03.ibm.com/systems/support/storage/config/ssic/displayesssearchwithoutjs.wss?start_over=yes

- Direct links to IBM tape and disk storage Web sites

  http://www-03.ibm.com/systems/storage/disk/
  http://www-03.ibm.com/systems/storage/tape/index.html

- IBM pre-requisites Web site (hardware and release level supporting selected hardware)

  http://www-912.ibm.com/e_dir/eServerPrereq.nsf/

- IBM planning Web site for System i

  http://www-304.ibm.com/systems/support/i/planning

- IBM planning, upgrades Web site for System i

  http://www-304.ibm.com/systems/support/i/planning/migrationupgrade.html

- Power Systems Facts and Figures Web site that includes matrixes of hardware feature and systems supported

  http://www-03.ibm.com/systems/p/hardware/reports/factsfeatures.html

- Performance management Web site information, starting at:

  http://www-03.ibm.com/systems/i/advantages/perfmgmt/resource.html

- Relative performance (rPerf) for AIX information

  http://www.ibm.com/systems/power/hardware/notices/rperf.html
  http://www-912.ibm.com/e_dir/eServerPrereq.nsf/

- News on new computer technologies

  http://www.ibm.com/chips/micronews

- IBM Smart Business pre-packaged solutions

  – IBM Business Partner - PartnerWorld (ISV and IBM Blog)

  https://www.ibm.com/communities/service/html/communityview?communityUuid=5c1dae47-e576-4601-9f01-6e13bb7ab027

  – IBM Press releases

  http://www-03.ibm.com/press/us/en/pressrelease/26091.wss

How to get Redbooks publications

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

ibm.com/redbooks

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services
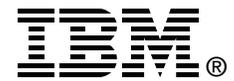
IBM Power 520 Technical Overview

Redpaper™

Covering the 8203-E4A supporting the AIX, i, and Linux operating systems

POWER6 processor-based PowerVM virtualization technology

Delivers IBM POWER6 efficiency

This IBM Redpaper publication is a comprehensive guide covering the IBM Power 520 server, machine type model 8203-E4A. The goal of this paper is to introduce this innovative server that includes IBM System i and IBM System p and new hardware technologies. The major hardware offerings include:

► The POWER6 processor, available at frequencies of 4.2 GHz and 4.7 GHz.
► Specialized POWER6 DDR2 memory that provides greater bandwidth, capacity, and reliability.
► The 1 Gb or 10 Gb Integrated Virtual Ethernet adapter that brings native hardware virtualization to this server.
► EnergyScale technology that provides features such as power trending, power-saving, capping of power, and thermal measurement.
► PowerVM virtualization technology.
► Mainframe continuous availability brought to the entry server environment.

This Redpaper expands the current set of IBM Power System documentation by providing a desktop reference that offers a detailed technical description of the Power 520 system.

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

REDP-4403-00