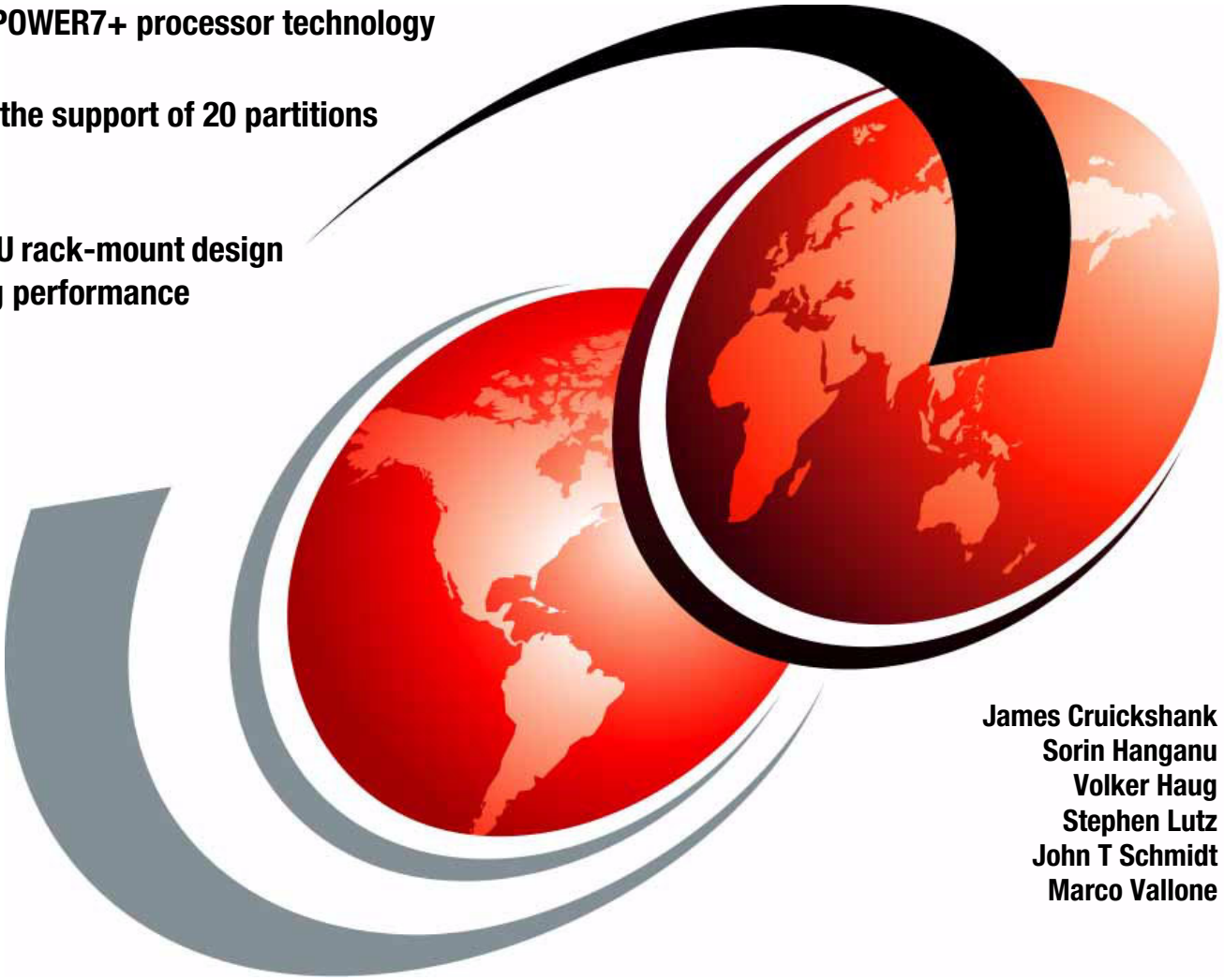
IBM

IBM Power 710 and 730 Technical Overview and Introduction

- Features 8231-E1D, 8231-E2D, 8246 PowerLinux servers based on POWER7+ processor technology
- Describes the support of 20 partitions per core
- Explains 2U rack-mount design for leading performance



James Cruickshank
Sorin Hanganu
Volker Haug
Stephen Lutz
John T Schmidt
Marco Vallone

Redpaper

ibm.com/redbooks

International Technical Support Organization

IBM Power 710 and 730 Technical Overview and Introduction

May 2013

REDP-4983-00

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

First Edition (May 2013)

This edition applies to the IBM Power 710 (8231-E1D) and Power 730 (8231-E2D), IBM PowerLinux 7R1 (8246-L1D and 8246-L1T), and PowerLinux 7R2 (8246-L2D and 826-L2T) servers.

© Copyright International Business Machines Corporation 2013. All rights reserved.
Note to U.S. Government Users Restricted Rights -- Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.

Contents

© Copyright IBM Corp. 2013. All rights reserved.

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 grant 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 websites are provided for convenience only and do not in any manner serve as an endorsement of those websites. The materials at those websites are not part of the materials for this IBM product and use of those websites 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 measurements 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.

© Copyright IBM Corp. 2013. All rights reserved.

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:

Active Memory™
AIX®
BladeCenter®
DS8000®
Dynamic Infrastructure®
Electronic Service Agent™
EnergyScale™
Focal Point™
IBM®
IBM Flex System™
IBM SmartCloud®
IBM Systems Director Active Energy Manager™
Micro-Partitioning®
POWER®
POWER Hypervisor™
Power Systems™
Power Systems Software™
POWER6®
POWER6+™
POWER7®
POWER7 Systems™
POWER7+™
PowerHA®
PowerLinux™
PowerPC®
PowerVM®
pSeries®
PureFlex™
Rational®
Rational Team Concert™
Real-time Compression™
Redbooks®
Redpaper™
Redpapers™
Redbooks (logo) ®
Storwize®
System p®
System Storage®
System x®
System z®
Tivoli®
XIV®

The following terms are trademarks of other companies:

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

ITIL is a registered trademark, and a registered community trademark of The Minister for the Cabinet Office, and is registered in the U.S. Patent and Trademark Office.

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

LTO, Ultrium, the LTO Logo and the Ultrium logo are trademarks of HP, IBM Corp. and Quantum in the U.S. and other countries.

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

Java, and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

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

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 710 (8231-E1D) and Power 730 (8231-E2D) servers that support IBM AIX®, IBM i, and Linux operating systems. This paper also describes the IBM PowerLinux™ 7R1 (8246-L1D and 8246-L1T) and the PowerLinux 7R2 (8246-L2D and 8246-L2T) servers that support the Linux operating system. The goal of this paper is to introduce the innovative Power 710, Power 730, PowerLinux 7R1, and PowerLinux 7R2 offerings and their major functions:

- The IBM POWER7+™ processor available at frequencies of 3.6 GHz, 4.2 GHz, and 4.3 GHz.
- The larger IBM POWER7+ Level 3 cache provides greater bandwidth, capacity, and reliability.
- The 4-port 10/100/1000 Base-TX Ethernet PCI Express adapter included in the base configuration and installed in a PCIe Gen2 x4 slot.
- The integrated SAS/SATA controller for HDD, SSD, tape, and DVD. This controller supports built-in hardware RAID 0, 1, and 10.
- New IBM PowerVM® V2.2.2 features, such as 20 LPARs per core.
- The improved IBM Active Memory™ Expansion technology that provides more usable memory than is physically installed in the system.
- IBM EnergyScale™ technology that provides features such as power trending, power-saving, capping of power, and thermal measurement.
- High-performance SSD drawer.

Professionals who want to acquire a better understanding of IBM Power Systems™ products can benefit from reading this publication. The intended audience includes these roles:

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

This paper complements the available set of IBM Power Systems documentation by providing a desktop reference that offers a detailed technical description of the Power 710 and Power 730 systems.

This paper does not replace the latest marketing materials and configuration 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.

© Copyright IBM Corp. 2013. All rights reserved.

Authors

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

James Cruickshank works on the Power Systems Client Technical Specialist team for IBM in the UK. He holds an Honors degree in Mathematics from the University of Leeds. James has over 11 years of experience working with IBM pSeries®, IBM System p®, and Power Systems products and is a member of the EMEA Power Champions team. James supports customers in the financial services sector in the UK.

Sorin Hanganu is an Accredited Product Services professional. He has eight years of experience working on Power Systems and IBM i products. He is an IBM Certified Solution Expert for IBM Dynamic Infrastructure® and also an IBM Certified Systems Expert for Power Systems, AIX, PowerVM virtualization, ITIL, and ITSM. Sorin works as a System Services Representative for Power Systems in Bucharest, Romania.

Volker Haug is an Open Group Certified IT Specialist within IBM Germany, supporting Power Systems clients and Business Partners as a Client Technical Specialist. He holds a diploma degree in Business Management from the University of Applied Studies in Stuttgart. His career includes more than 25 years of experience with Power Systems, AIX, and PowerVM virtualization; he has written several Power Systems and PowerVM IBM Redbooks® publications. Volker is an IBM POWER7® Champion and a member of the German Technical Expert Council, an affiliate of the IBM Academy of Technology.

Stephen Lutz is a Certified Senior Technical Sales Professional for Power Systems, working for IBM Germany. He holds a degree in Commercial Information Technology from the University of Applied Science Karlsruhe, Germany. He is POWER7 champion and has 14 years experience in AIX, Linux, virtualization, and Power Systems and its predecessors, providing pre-sales technical support to clients, Business Partners, and IBM sales representatives all over Germany. Stephen is also an expert in IBM Systems Director, its plug-ins, and IBM SmartCloud® Entry with a focus on Power Systems and AIX.

John T Schmidt is an Accredited IT Specialist for IBM and has 12 years of experience with IBM and Power Systems. He has a degree in Electrical Engineering from the University of Missouri - Rolla, and an MBA from Washington University in St. Louis. In addition to contributing to eight other Power Systems IBM Redpapers™ publications, in 2010, he completed an assignment with the IBM Corporate Service Corps in Hyderabad, India. He is working in the United States as a pre-sales Field Technical Sales Specialist for Power Systems in Boston, MA.

Marco Vallone is a Certified IT Specialist at IBM, Italy. He joined IBM in 1989, starting in the Power Systems production plant (Santa Palomba) as a Product Engineer, and then worked for the ITS AIX support and delivery service center. For the last eight years of his career, he has worked as IT Solution Architect in the ITS Solution Design Compentence Center of Excellence in Rome, where he mainly designs infrastructure solutions on distributed environments with a special focus on Power System solutions.

The project that produced this publication was managed by:

Scott Vetter
Executive Project Manager, PMP

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

Larry L. Amy, Ron Arroyo, Hsien-I Chang, Carlo Costantini, Kirk Dietzman, Gary Elliott, Michael S. Floyd, James Hermes, Pete Heyrman, John Hilburn, Roberto Huerta de la Torre, Dan Hurlimann, Roxette Johnson, Sabine Jordan, Kevin Kehne, Robert Lowden, Jia Lei Ma, Hilary Melville, Hans Mozes, Thoi Nguyen, Mark Olson, Robb Romans, Pat O'Rourke, Jan Palmer, Velma Pavlasek, Dave Randall, Todd Rosedahl, Jeff Stuecheli, Madeline Vega
IBM

Udo Sachs
SVA Germany

Tamikia Barrow
International Technical Support Organization, Austin Center

Now you can become a published author, too!

Here's an opportunity to spotlight your skills, grow your career, and become a published author—all at the same time! Join an ITSO residency project and help write a book in your area of expertise, while honing your experience using leading-edge technologies. Your efforts will help to increase product acceptance and customer satisfaction, as you expand your network of technical contacts and relationships. Residencies run from two to six weeks in length, and you can participate either in person or as a remote resident working from your home base.

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 publications in one of the following ways:

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

  ibm.com/redbooks

- Send your comments in an email 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

Stay connected to IBM Redbooks

- Find us on Facebook:

  http://www.facebook.com/IBMRedbooks

- Follow us on Twitter:

  http://twitter.com/ibmredbooks

- Look for us on LinkedIn:

  http://www.linkedin.com/groups?home=&gid=2130806

- Explore new Redbooks publications, residencies, and workshops with the IBM Redbooks weekly newsletter:

  https://www.redbooks.ibm.com/Redbooks.nsf/subscribe?OpenForm

- Stay current on recent Redbooks publications with RSS Feeds:

  http://www.redbooks.ibm.com/rss.html

1

General description

The IBM Power 710 (8231-E1D) and IBM Power 730 servers (8231-E2D) use the latest POWER7+ processor technology that is designed to deliver unprecedented performance, scalability, reliability, and manageability for demanding commercial workloads.

The high data transfer rates that are offered by the Peripheral Component Interconnect Express (PCIe) Gen2 slots can allow higher I/O performance or consolidation of the I/O demands on to fewer adapters running at higher rates. The result is better system performance at a lower cost when I/O demands are high.

The Power 710 server is a high-performance, energy-efficient, reliable, and secure infrastructure and application server in a dense form factor. It contains innovative workload-optimizing technologies that maximize performance, based on client computing needs and intelligent energy features that help maximize performance and optimize energy-efficiency. The result is one of the most cost-efficient solutions for AIX, IBM i, and Linux deployments.

The IBM Power 730 server delivers the outstanding performance of the POWER7+ processor in a dense, rack-optimized form factor and is ideal for running multiple application and infrastructure workloads in a virtualized environment. You can take advantage of the Power 730 server's scalability and capacity by using the IBM industrial strength PowerVM technology to fully employ the server's capability.

© Copyright IBM Corp. 2013. All rights reserved.

1.1 Systems overview

The following sections provide detailed information about the Power 710 and Power 730 systems.

1.1.1 The Power 710 server

The IBM Power 710 server is a 2U rack-mount server with one processor socket, offering 4-core 3.6 GHz, 6-core 4.2 GHz, and 8-core 4.2-GHz configurations. The POWER7+ processor chips in this server are 64-bit, available as 4-core, 6-core, and 8-core modules with 10 MB of L3 cache per core and 256 KB of L2 cache per core.

The Power 710 server supports a maximum of eight DDR3 DIMM slots, with four DIMM slots included in the base configuration, and four DIMM slots available with an optional memory riser card. This configuration allows for a maximum system memory of 256 GB.

The POWER7+ chip includes a hardware accelerator for Active Memory Expansion. This accelerator provides 25% higher levels of memory expansion than available with POWER7 chips. While IBM POWER7 Systems™ offer up to 100% memory expansion, which can effectively double the server's maximum memory, POWER7+ servers offer up to 125% memory expansion for AIX partitions. Thus a system memory maximum of 256 GB can effectively become more than 512 GB effective memory capacity.

The Power 710 server offers three storage backplane options. The first supports three SFF SAS HDDs or SSDs, a SATA DVD, and a half-high tape drive. The second supports six SFF SAS HDDs or SSDs and a SATA DVD. These choices both provide an integrated SAS controller offering RAID 0, 1, and 10 support. The third option supports six SFF SAS HDDs or SSDs, a SATA DVD, and adds support for Dual Write Cache RAID 5 or 6, and an external SAS port. HDDs and SSDs are hot-swap and front accessible with each of the alternatives.

The Power 710 includes five PCI Express (PCIe) Gen2 low profile (LP) slots for installing adapters in the system. The system also includes a PCIe x4 Gen2 Low Profile expansion slot that contains a 4-Port 10/100/1000 Base-TX Ethernet PCI Express Gen2 adapter.

Remember: The Integrated Virtual Ethernet (IVE) adapter is not available for the Power 710.

Figure 1-1 shows the Power 710 server that contains six SFF disk drives and a DVD drive.

Figure 1-1 IBM Power 710 server

1.1.2 The Power 730 server

The IBM Power 730 server is a 2U rack-mount server with two processor sockets that offers 8-core 4.3 GHz, 12-core 4.2 GHz, and 16-core 3.6 GHz and 4.2 GHz configurations. The POWER7+ processor chips in this server are 64-bit, available as 4-core, 6-core, and 8-core modules with 10 MB of L3 cache per core and 256 KB of L2 cache per core. The new Power 730 also provides expanded I/O capabilities that use the high-performance Gen2 PCIe interfaces, and includes the capability of additional I/O using the 12x PCIe I/O expansion drawers.

The Power 730 server supports a maximum of 16 DDR3 DIMM slots, with four DIMM slots included in the base configuration. A maximum of three additional memory riser cards, each containing four DIMM slots, allowing a maximum system memory of 512 GB.

The POWER7+ chip includes a hardware accelerator for Active Memory Expansion. This accelerator provides 25% higher levels of memory expansion than available with POWER7 chips. While POWER7 Systems offer up to 100% memory expansion, which can effectively double the server's maximum memory, POWER7+ servers offer up to 125% memory expansion for AIX partitions. Thus a system memory maximum of 512 GB can effectively become more than 1024 GB effective memory capacity.

The Power 730 server offers three storage backplane options. The first supports three SFF SAS hard disk drives (HDDs) or solid-state drives (SSDs), a SATA DVD, and a half-high tape drive. The second supports six SFF SAS HDDs or SSDs and a SATA DVD. These choices both provide an integrated SAS controller, offering RAID 0, 1, and 10 support. The third option supports six SFF SAS HDDs or SSDs, a SATA DVD, and adds support for Dual Write Cache RAID 5, 6, and an external SAS port. HDDs and SSDs are hot-swap and front accessible with each of the alternatives.

The Power 730 includes five PCI Express (PCIe) Gen2 low profile (LP) slots for installing adapters in the system. The system also includes a PCIe x4 Gen2 low profile expansion slot that contains a 4-Port 10/100/1000 Base-TX Ethernet PCI Express adapter.

Remember: The Integrated Virtual Ethernet (IVE) adapter is not available for the Power 730.

Figure 1-2 shows the Power 730 server that contains three SFF disk drives, a DVD drive, and a tape drive.

Figure 1-2 IBM Power 730

1.1.3 IBM PowerLinux 7R1 server

The IBM PowerLinux 7R1 (8246-L1D and 8246-L1T) server delivers the outstanding performance of the IBM POWER7+ processor in a dense, highly efficient 2U rack-optimized form factor for Linux clients. It is ideal for running multiple Linux infrastructure and application workloads, and with PowerVM virtualization, can be more economical than traditional Linux servers.

Take advantage of the scalability and capacity of IBM PowerLinux 7R1 by using the feature-rich PowerVM virtualization technology from IBM to fully use the server's capacity and deploy virtual partitions faster. You can move workloads as needed across PowerLinux and Power Systems servers with Live Partition Mobility.

The IBM PowerLinux 7R1 server is a Linux only 2U rack-mount server with one processor socket offering 4-core 3.6 GHz, 6-core 4.2 GHz, and 8-core 4.2 GHz configurations. The new PowerLinux 7R1(8246-L1T only) server also delivers expanded storage capabilities by using high-performance Gen2 PCIe interfaces and more disk drives by using the EXP24S small form factor (SFF) Gen2 bay drawer.

The PowerLinux 7R1 server supports a maximum of eight DDR3 DIMM slots, with four DIMM slots included in the base configuration and four more slots available with one optional memory riser card, allowing for a maximum system memory of 256 GB.

Supported memory features (two memory DIMMs per feature) are 8 GB, 16 GB, 32 GB, and 64 GB running at speeds of 1066 MHz. PowerVM now features Active Memory Sharing, the technology you can use to intelligently exchange memory between running partitions for increased optimization of physical memory resources. Active Memory Sharing enables the sharing of a pool of physical memory among logical partitions (LPARs) on a single server, helping to increase memory utilization and drive down system costs.

The IBM PowerLinux 7R1 server offers three storage backplane options. The first supports three SFF SAS HDDs or SSDs, a SATA DVD, and a half- high tape drive. The second supports six SFF SAS HDDs or SSDs and a SATA DVD. These two choices both provide an integrated SAS controller, offering RAID 0, 1, and 10 support. The third supports six SFF SAS HDDs or SSDs and a SATA DVD, and adds support for dual write cache RAID 5, RAID 6, and an external SAS port. HDDs and SSDs are hot-swap and front-accessible with each of the three alternatives.

Other integrated features include the following items:

- Five PCIe x8 Gen2 low profile expansion slots.
- PCIe2 LP 4-Port 1 Gb Ethernet adapter.
- Service processor.
- Integrated SAS and SATA controller for HDD, SSD, tape, and DVD in the system unit, supporting RAID 0, 1, and 10; RAID 5 and RAID 6 also available.
- EnergyScale technology.
- Two system ports, three USB ports, and two HMC ports.
- One 1925 watt AC power supply is required. A second power supply is available for redundant power.
- Redundant and hot-swap cooling.

The PowerLinux system is specifically designed for emerging workloads that are proven ideal for a virtualized scale-out, Linux server environment.

The PowerLinux 7R1 server benefits from POWER7+ performance, Intelligent Threads technology, and high memory and I/O bandwidth. These workloads realize more performance, more efficient virtualization, unique workload optimizing features, and industry-leading reliability, availability, and scalability at prices comparable with traditional Linux servers.

The PowerLinux 7R1 is based on the Power 710 server. The firmware on the PowerLinux 7R1 is modified to allow only Virtual I/O Server (VIOS) and Linux operating systems to run in LPARs. Most features of the Power 710 as described in this paper also apply to the PowerLinux 7R1 even when the PowerLinux 7R1 is not mentioned explicitly. Where differences exist between the Power 710 and Power 7R1, this paper highlights them.

L1D and L1T differences: The 8246-L1T supports connection to external disk drawers. The 8246-L1D does not offer connection to external I/O drawers.

Figure 1-3 shows a PowerLinux 7R1 server containing three SFF disk drives, a DVD drive, and a tape drive.

Figure 1-3 IBM PowerLinux 7R1

1.1.4 IBM PowerLinux 7R2 server

The IBM PowerLinux 7R2 (8246-L2D and 8246-L2T) server delivers the outstanding performance of the IBM POWER7+ processor in a dense, highly efficient 2U rack-optimized form factor for Linux clients. It is ideal for running multiple Linux infrastructure and application workloads, virtualized with PowerVM, more economically than traditional Linux servers. Take advantage of the scalability and capacity of the IBM PowerLinux 7R2 with the IBM feature-rich PowerVM virtualization technology to fully use the server's capacity and deploy virtual partitions faster. You can move workloads, as needed, across PowerLinux and Power Systems servers with Live Partition Mobility.

The PowerLinux 7R2 server is a Linux-only 2U rack-mount server with two processor sockets, offering 16-core 3.6 GHz and 4.2 GHz POWER7+ configurations. The new PowerLinux 7R2 (8246-L2T only) server also provides expanded I/O capabilities with the high-performance Gen2 PCIe interfaces, and includes the capability of additional I/O that use the 12x PCIe I/O expansion drawers.

The PowerLinux 7R2 server supports a maximum of 16 DDR3 DIMM slots, with four DIMM slots included in the base configuration, and 12 DIMM slots available with three optional memory riser cards, allowing for a maximum system memory of 512 GB.

Supported memory features (two memory DIMMs per feature) are 8 GB, 16 GB, 32 GB, and 64 GB, and run speeds of 1066 MHz. Also, PowerVM now features Active Memory Sharing, the technology that you can use to intelligently exchange memory between running partitions for increased optimization of physical memory resources. Active Memory Sharing enables the sharing of a pool of physical memory among logical partitions (LPARs) on a single server, helping to increase memory utilization and drive down system costs.

The PowerLinux 7R2 server offers three storage backplane options. The first supports three SFF SAS HDDs or SSDs, a SATA DVD, and a half-high tape drive. The second supports six SFF SAS HDDs or SSDs and a SATA DVD. These two choices both provide an integrated SAS controller, offering RAID 0, 1, and 10 support. The third supports six SFF SAS HDDs or SSDs, and a SATA DVD, and adds support for Dual Write Cache RAID 5, RAID 6, and an external SAS port. HDDs and SSDs are hot-swap and front accessible with each of the three alternatives.

Other integrated features include the following items:

- Five PCIe x8 Gen2 low profile expansion slots.
- PCIe2 LP 4-Port 1 Gb Ethernet adapter.
- Two GX++ slots for 12X I/O loop.
- Service processor.
- Integrated SAS and SATA controller for HDD, SSD, tape, and DVD in the system unit, supporting RAID 0, 1, and 10. RAID 5 and RAID 6 are available.
- EnergyScale technology.
- Two system ports, three USB ports, and two HMC ports.
- Redundant and hot-swap power.
- Redundant and hot-swap cooling.

The PowerLinux system is specifically designed for emerging workloads that are proven ideal for a virtualized scale-out, Linux server environment. The PowerLinux 7R2 server benefits from POWER7+ performance, Intelligent Threads technology, and high memory and I/O bandwidth. These workloads benefit from POWER7+ processor performance, virtualization efficiencies, unique workload optimizing features, and industry-leading reliability, availability, and scalability at prices comparable with traditional Linux servers.

The PowerLinux 7R2 is based on the Power 730 server. The firmware on the PowerLinux 7R2 is modified to allow only VIOS and Linux operating systems to run in LPARs. Most features of the Power 730 as described in this paper also apply to the PowerLinux 7R2, even when the PowerLinux 7R2 is not mentioned explicitly. Where differences exist between the Power 730 and Power 7R2, this paper attempts to highlight them.

L2D and L2T differences: The 8246-L2T supports connection to external I/O drawers. The 8246-L2D does not offer connection to external I/O drawers.

Figure 1-4 shows the IBM PowerLinux 7R2 system that contains six SFF drives and a DVD drive.

Figure 1-4 IBM PowerLinux 7R2

1.2 Operating environment

Table 1-1 lists the operating environment specifications for the servers.

Table 1-1 Operating environment for Power 710 and Power 730

Power 710 and Power 730 operating environment				
Description	**Operating**		**Non-operating**	
	Power 710	**Power 730**	**Power 710**	**Power 730**
Temperature	5 - 40[a] degrees C (41 - 104 degrees F) Recommended: 18 to 27 degrees C (64 to 80 degrees F)		5 - 45 degrees C (41 - 113 degrees F)	
Relative humidity	8 - 80%		8 - 80%	
Maximum dew point	28 degrees C (84 degrees F)		N/A	
Operating voltage	100 - 127 V AC or 200 - 208 V AC or 200 - 240 V AC		N/A	
Operating frequency	50 or 60 Hz		N/A	
Power consumption	925 Watts maximum	1,368 Watts maximum	N/A	
Power source loading	0.944 kVA maximum	1.396 kVA maximum	N/A	
Thermal output	3156 BTU/hour maximum	4668 BTU/hour maximum	N/A	
Maximum altitude	3,050 m (10,000 ft)		N/A	

a. Heavy workloads may see some performance degradation above 35C if internal temperatures trigger a CPU clock reduction.

Tip: The maximum measured value is expected from a fully populated server under an intensive workload. The maximum measured value also accounts for component tolerance and operating conditions that are not ideal. Power consumption and heat load vary greatly by server configuration and utilization. Use the IBM Systems Energy Estimator to obtain a heat output estimate based on a specific configuration, available at the following website:

http://www-912.ibm.com/see/EnergyEstimator

1.3 Physical package

Table 1-2 shows the physical dimensions of the Power 710 and Power 730 chassis. Both servers are available only in a rack-mounted form factor. Each takes 2U (2 EIA units) of rack space.

Table 1-2 Physical dimensions

Dimension	Power 710 (8231-E1D)	Power 730 (8231-E2D)
Width	440 mm (19.0 in)	440 mm (19.0 in)
Depth	706 mm (27.8 in)	706 mm (27.8 in)
Height	89 mm (3.5 in)	89 mm (3.5 in)
Weight (maximum configuration)	28.2 kg (62 lbs)	29.5 kg (62 lbs)

Figure 1-5 shows the rear view of a Power 730 system.

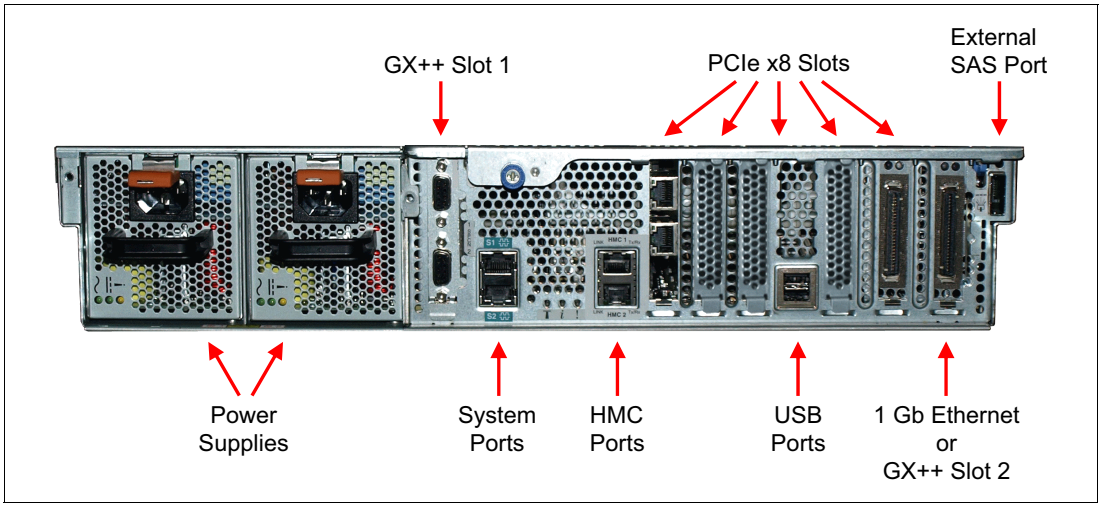


Figure 1-5 Rear view of a Power 730 system

1.4 System features

The system chassis contains one processor module (Power 710) or two processor modules (Power 730). Each POWER7+ processor module is either 4-core, 6-core, or 8-core. Each of the POWER7+ processors in the server has a 64-bit architecture, up to 2 MB of L2 cache (256 KB per core), and up to 80 MB of L3 cache (10 MB per core).

1.4.1 Power 710 system features

This summary describes the standard features of the Power 710:

- Rack-mount (2U) chassis
- Single processor module:
  - 4-core 3.6 GHz processor module
  - 6-core 4.2 GHz processor module
  - 8-core 4.2 GHz processor module
- Up to 256 GB of 1066 MHz DDR3 ECC memory
- Choice of three disk and media backplanes:
  - Six 2.5-inch SAS HDDs or SSDs and one DVD bay, and an integrated SAS controller, offering RAID 0, 1, and 10 support
  - Six 2.5-inch SAS HDDs or SSDs, an SATA DVD, and adds support for Dual Write Cache RAID 5, 6, and an external SAS port
  - Three 2.5-inch HDD/SSD/Media backplane with one tape drive bay and one DVD bay, an integrated SAS controller, offering RAID 0, 1, and 10 support
- A PCIe x4 Gen2 Low Profile expansion slot containing a 4-Port 10/100/1000 Base-TX Ethernet PCI Express adapter
- Five PCIe Gen2 x8 low profile slots
- One GX++ slot
- Integrated features:
  - Service processor
  - EnergyScale technology
  - Hot-swap and redundant cooling fans
  - Three USB ports
  - Two system ports
  - Two HMC ports
- Optional redundant, 1925 Watt AC hot-swap power supplies

1.4.2 Power 730 system features

This summary describes the standard features of the Power 730:

- Rack-mount (2U) chassis
- Two processor modules:
  - 8-core configuration using two 4-core 4.3 GHz processor modules
  - 12-core configuration using two 6-core 4.2 GHz processor modules
  - 16-core configuration using two 8-core 3.6 GHz processor modules
  - 16-core configuration using two 8-core 4.2 GHz processor modules
- Up to 512 GB of 1066 MHz DDR3 ECC memory
- Choice of three disk/media backplanes:
  - Six 2.5-inch SAS HDDs or SSDs and one DVD bay, and an integrated SAS controller, offering RAID 0, 1, and 10 support
  - Six 2.5-inch SAS HDDs or SSDs, an SATA DVD, and adds support for Dual Write Cache RAID 5, 6, and an external SAS port

  - Three 2.5-inch HDD/SSD/Media backplane with one tape drive bay and one DVD bay, and an integrated SAS controller, offering RAID 0, 1, and 10 support
- A PCIe x4 Gen2 Low Profile expansion slot with either a 4-Port 10/100/1000 Base-TX Ethernet PCI Express adapter or a GX++ Dual-Port 12x Channel Attach adapter
- Five PCIe Gen2 x8 low profile slots
- Two GX++ slots
- Integrated features:
  - Service processor
  - EnergyScale technology
  - Hot-swap and redundant cooling fans
  - Three USB ports
  - Two system ports
  - Two HMC ports
- Two redundant power supplies, 1925 Watt AC, hot-swap

PowerLinux 7R2: The PowerLinux 7R2 only supports 16-core configurations. The 8-core and 12-core configurations are not supported.

1.4.3 Minimum features

Each system has a minimum feature set to be a valid configuration.

The minimum Power 710 initial order must include a processor module, processor activations, memory, one HDD/SSD, a storage backplane, a power supply and power cord, an operating system indicator, a chassis indicator, and a language-group specify.

The minimum Power 730 initial order must include two processor modules, processor activations, memory, one HDD/SSD, a storage backplane, two power supplies and power cords, an operating system indicator, a chassis indicator, and a language group specify.

If IBM i is the primary operating system (FC 2145), the initial order must also include one additional HDD or SSD, mirrored system disk-level specify code, and a system console indicator. A DVD is defaulted on every order but can be deselected. A DVD-ROM or DVD-RAM must be accessible by the system.

Boot from SAN option: If AIX, IBM i, or Linux is the primary operating system, no internal HDD or SSD is required if feature SAN Load Source Specify (Boot from SAN), FC 0837, is selected. A Fibre Channel or FCoE adapter must be ordered if FC 0837 is selected.

1.4.4 Power supply features

One 1925 watt AC power supply (FC 5532) is required for the Power 710. A second power supply is optional. Two 1925 watt AC power supplies (FC 5532) are required for the Power 730. The second power supply provides redundant power for enhanced system availability. To provide full redundancy, the two power supplies must be connected to separate power sources.

The server will continue to function with one working power supply. A failed power supply can be hot-swapped but must remain in the system until the replacement power supply is available for exchange.

1.4.5 Processor module features

Each processor module in the system houses a single POWER7+ processor chip. The processor chip has either four cores, six cores, or eight cores. The Power 710 supports one processor module. The Power 730 supports two processor modules. Both processor modules in the system must be identical.

The number of installed cores in a Power 710 or Power 730 must be equal to the number of ordered activation features. Cells marked N/A indicate bulk ordering codes, and Custom Card Identification Number (CCIN) is not applicable. A blank CCIN cell indicates that CCIN is not available.

Table 1-3 summarizes the processor features that are available for the Power 710.

Table 1-3 Processor features for the Power 710

Feature code	Processor module description
EPCE	4-core 3.6 GHz POWER7+ processor module
EPCG	6-core 4.2 GHz POWER7+ processor module
EPCJ	8-core 4.2 GHz POWER7+ processor module

The Power 730 requires that two identical processor modules be installed. Table 1-4 lists the available processor features.

Table 1-4 Processor features for the Power 730

Feature code	Processor module description
EPCF	4-core 4.3 GHz POWER7+ processor module
EPCG	6-core 4.2 GHz POWER7+ processor module
EPCH	8-core 3.6 GHz POWER7+ processor module
EPCJ	8-core 4.2 GHz POWER7+ processor module

1.4.6 Memory features

In POWER7+ processor-based systems, DDR3 memory is used throughout. The POWER7+ DDR3 memory uses a new memory architecture to provide greater bandwidth and capacity. This architecture enables operating at a higher data rate for larger memory configurations.

Memory in the systems is installed into memory riser cards. One memory riser card is included in the base system. The base memory riser card does not appear as a feature code (FC) in the configurator. One additional memory riser card, FC 5265, can be installed in the Power 710. Three additional memory riser cards, FC 5265, can be installed in the Power 730. The FC 5265 is replaced with FC EL0A on PowerLinux systems. Each memory riser card provides four DDR3 DIMM slots. DIMMs are available in capacities of 4 GB, 8 GB, 16 GB, and 32 GB at 1066 MHz and are installed in pairs.

Table 1-5 lists memory features that are available on the Power 710 and Power 730 systems.

Table 1-5 Summary of memory features

Feature code	Feature capacity	Access rate	DIMMs
EM08	8 GB	1066 MHz	2 x 4 GB DIMMs
EM4B	16 GB	1066 MHz	2 x 8 GB DIMMs
EM4C	32 GB	1066 MHz	2 x 16 GB DIMMs
EM4D	64 GB	1066 MHz	2 x 32 GB DIMMs

Generally, the best approach is for memory to be installed evenly across all memory riser cards in the system. Balancing memory across the installed memory riser cards allows memory access in a consistent manner and typically results in the best possible performance for your configuration.

Remember: The memory cards operate at lower voltage to save energy. Therefore, they cannot be interchanged with the 8 GB and 16 GB memory features that are used within the 8231-E2B model.

1.5 Disk and media features

The Power 710 and Power 730 systems feature an integrated SAS controller, offering RAID 0, 1, and 10 support with three storage backplane options:

- FC EJ0D supports six SFF disk units, either HDD or SSD, and an SATA DVD. There is no support for split backplane and for RAID 5 or 6.
- FC EJ0E supports three small form-factor (SFF) disk units, either HDD or SSD, an SATA DVD, and a tape. There is no support for split backplane and for RAID 5 or 6.
- FC EJ0F supports six SFF disk units, either HDD or SSD, an SATA DVD, and an external SAS port. RAID 5 and 6 are supported.

Table 1-6 shows the available disk drive feature codes that can be installed in the Power 710.

Table 1-6 Disk drive feature code description

Feature code	CCIN	Description	OS support
1751		900 GB 10K RPM SAS SFF Disk Drive (AIX, Linux)	AIX, Linux
1752		900 GB 10K RPM SAS SFF-2 Disk Drive (AIX, Linux)	AIX, Linux
1775	58B3	177 GB SFF-1 SSD with eMLC for AIX, Linux	AIX, Linux
1790		600 GB 10K RPM SAS SFF Disk Drive (AIX, Linux)	AIX, Linux
1880	169C	300 GB 15K RPM SAS SFF Disk Drive (AIX, Linux)	AIX, Linux
1885		300 GB 10K RPM SFF SAS Disk Drive	AIX, Linux
1886		146 GB 15K RPM SFF SAS Disk Drive (AIX, Linux)	AIX, Linux
1917		146 GB 15K RPM SAS SFF-2 Disk Drive (AIX, Linux)	AIX, Linux
1925		300 GB 10K RPM SAS SFF-2 Disk Drive (AIX, Linux)	AIX, Linux

Feature code	CCIN	Description	OS support
1953		300 GB 15K RPM SAS SFF-2 Disk Drive (AIX, Linux)	AIX, Linux
1964		600 GB 10K RPM SAS SFF-2 Disk Drive (AIX, Linux)	AIX, Linux
ES02	58BB	387 GB 1.8" SAS SSD for AIX, Linux with eMLC	AIX, Linux
ES0A	58B8	387 GB SFF-1 SSD for AIX, Linux with eMLC	AIX, Linux
ES0C	58B9	387 GB SFF-2 SSD for AIX, Linux with eMLC	AIX, Linux
ESR2	N/A	Six FC ES02 387 GB 1.8" SAS SSD for AIX, Linux with eMLC	AIX, Linux
ESRA	N/A	Four FC ES0A 387 GB SFF-1 SSD for AIX, Linux with eMLC	AIX, Linux
ESRC	N/A	Four FC ES0C 387 GB SFF-2 SSD for AIX, Linux with eMLC	AIX, Linux
1737	19A4	856 GB 10K RPM SAS SFF Disk Drive (IBM i)	IBM i
1738	19B4	856 GB 10K RPM SAS SFF-2 Disk Drive (IBM i)	IBM i
1787	58B3	177 GB SFF-1 SSD with eMLC (IBM i)	IBM i
1879	19A1	283 GB 15K RPM SAS SFF Disk Drive (IBM i)	IBM i
1888	198C	139 GB 15K RPM SFF SAS Disk Drive (IBM i)	IBM i
1911	198D	283 GB 10K RPM SFF SAS Disk Drive (IBM i)	IBM i
1916	19A3	571 GB 10K RPM SAS SFF Disk Drive (IBM i)	IBM i
ES0B	58B8	387 GB SFF-1 SSD for IBM i with eMLC	IBM i
ES0D	58B9	387 GB SFF-2 SSD for IBM i with eMLC	IBM i
ESRB	N/A	Four FC ES0B 387 GB SFF-1 SSD for IBM i with eMLC	IBM i
ESRD	N/A	Four FC ES0D 387 GB SFF-2 SSD for IBM i with eMLC	IBM i

Table 1-7 shows the available disk drive feature codes that can be installed in the Power 730.

Table 1-7 Disk drive feature code description

Feature code	CCIN	Description	OS support
1751		900 GB 10K RPM SAS SFF Disk Drive (AIX, Linux)	AIX, Linux
1752		900 GB 10K RPM SAS SFF-2 Disk Drive (AIX, Linux)	AIX, Linux
1775	58B3	177 GB SFF-1 SSD with eMLC (AIX, Linux)	AIX, Linux
1790		600 GB 10K RPM SAS SFF Disk Drive (AIX, Linux)	AIX, Linux
1793	58B4	177 GB SFF-2 SSD with eMLC (AIX, Linux)	AIX, Linux
1866	N/A	Quantity 150 of FC 1917	AIX, Linux
1880	169C	300 GB 15K RPM SAS SFF Disk Drive (AIX, Linux)	AIX, Linux
1885		300 GB 10K RPM SFF SAS Disk Drive	AIX, Linux
1886		146 GB 15K RPM SFF SAS Disk Drive (AIX, Linux)	AIX, Linux
1887	N/A	Quantity 150 of FC 1793	AIX, Linux

Feature code	CCIN	Description	OS support
1917		146 GB 15K RPM SAS SFF-2 Disk Drive (AIX, Linux)	AIX, Linux
1925		300 GB 10K RPM SAS SFF-2 Disk Drive (AIX, Linux)	AIX, Linux
1953		300 GB 15K RPM SAS SFF-2 Disk Drive (AIX, Linux)	AIX, Linux
1964		600 GB 10K RPM SAS SFF-2 Disk Drive (AIX, Linux)	AIX, Linux
EQ0C	N/A	Quantity of 150 FC ES0C	AIX, Linux
EQ52	N/A	Quantity 150 of FC 1752 (900 GB SFF-2 disk)	AIX, Linux
ES02	58BB	387 GB 1.8" SAS SSD for AIX, Linux with eMLC	AIX, Linux
ES0A	58B8	387 GB SFF-1 SSD for AIX, Linux with eMLC	AIX, Linux
ES0C	58B9	387 GB SFF-2 SSD for AIX, Linux with eMLC	AIX, Linux
ESR2	N/A	Six FC ES02 387 GB 1.8" SAS SSD for AIX, Linux with eMLC	AIX, Linux
ESRA	N/A	Four FC ES0A 387 GB SFF-1 SSD for AIX, Linux with eMLC	AIX, Linux
ESRC	N/A	Four FC ES0C 387 GB SFF-2 SSD for AIX, Linux with eMLC	AIX, Linux
1737	19A4	856 GB 10K RPM SAS SFF Disk Drive (IBM i)	IBM i
1738	19B4	856 GB 10K RPM SAS SFF-2 Disk Drive (IBM i)	IBM i
1787	58B3	177 GB SFF-1 SSD with eMLC (IBM i)	IBM i
1794	58B4	177 GB SFF-2 SSD with eMLC (IBM i)	IBM i
1879	19A1	283 GB 15K RPM SAS SFF Disk Drive (IBM i)	IBM i
1888	198C	139 GB 15K RPM SFF SAS Disk Drive (IBM i)	IBM i
1911	198D	283 GB 10K RPM SFF SAS Disk Drive (IBM i)	IBM i
1916	19A3	571 GB 10K RPM SAS SFF Disk Drive (IBM i)	IBM i
1947	19B0	139 GB 15K RPM SAS SFF-2 Disk Drive (IBM i)	IBM i
1948	19B1	283 GB 15K RPM SAS SFF-2 Disk Drive (IBM i)	IBM i
1956	19B7	283 GB 10K RPM SAS SFF-2 Disk Drive (IBM i)	IBM i
1958	N/A	Quantity 150 of FC 1794	IBM i
1962	19B3	571 GB 10K RPM SAS SFF-2 Disk Drive (IBM i)	IBM i
EQ0D	N/A	Quantity of 150 FC ES0D	IBM i
ES04		387 GB 1.8" SAS SSD for IBM i with eMLC	IBM i
ES0B	58B8	387 GB SFF-1 SSD for IBM i with eMLC	IBM i
ES0D	58B9	387 GB SFF-2 SSD for IBM i with eMLC	IBM i
ESR4	N/A	Six FC ES04 387 GB 1.8" SAS SSD for IBM i with eMLC	IBM i
ESRB	N/A	Four FC ES0B 387 GB SFF-1 SSD for IBM i with eMLC	IBM i
ESRD	N/A	Four FC ES0D 387 GB SFF-2 SSD for IBM i with eMLC	IBM i

Tip: Be aware of the following considerations for SAS bay-based SSDs (FC 1775, FC 1787, FC 1793, FC 1794, FC 1890, and FC 1909):

- SFF features ES0A, ES0B, ESRA, ESRB, 1775, and 1787 are supported in the Power 710 and Power 730 system enclosure.
- FC 1793 and FC 1794 are not supported in the Power 710 and Power 730 system enclosure. They are supported only in the EXP24S SFF Gen2-bay Drawer (FC 5887).
- SSDs and disk drives (HDDs) are not allowed to mirror each other.
- When an SSD is placed in the FC EJ0F backplane, no EXP12S Expansion Drawer (FC 5886) or EXP24S SFF Gen2-bay Drawer (FC 5887) is supported to connect to the external SAS port.
- Regarding HDD/SSD data protection, if IBM i (FC 2145) is selected, one of the following items is required:
  - Disk mirroring (default), which requires feature code FC 0040, or FC 0308
  - SAN boot (FC 0837)
  - RAID, which requires FC EJ0F
  - Mixed Data Protection (FC 0296)

If you need more disks than are available with the internal disk bays, you can attach additional external disk subsystems. For more information about the available external disk subsystems, see 2.10, “External disk subsystems” on page 78.

SCSI disks are not supported in the Power 710 and Power 730 disk bays. Also, because no PCIe LP SCSI adapter is available, you cannot attach existing SCSI disk subsystems.

Split backplane: The Power 710 and Power 730 models do not support the split backplane function.

The Power 710 and Power 730 have a slim media bay that can contain an optional DVD-RAM (FC 5762). If FC EJ0E is selected for the storage backplane, a tape bay is available that can contain a tape drive or removable disk drive.

Internal docking station: The Internal Docking Station for Removable Disk Drive (FC EU23) is supported for AIX and Linux.

Table 1-8 shows the available media device feature codes for Power 710 and Power 730.

Table 1-8 Media device feature code description for Power 710 and 730

Feature code	Description
5762	SATA Slimline DVD-RAM Drive
EU23	RDX USB Internal Docking Station for Removable Disk Cartridge
EU04	RDX USB External Docking Station for Removable Disk Cartridge

For more information about the internal disk features, see 2.8, “Internal storage” on page 68.

1.6 I/O drawers for Power 710 and Power 730 servers

The Power 710 and Power 730 servers support the attachment of I/O drawers. The Power 710 supports disk-only I/O drawers (FC 5886, FC 5887, and FC EDR1), providing large storage capacity and multiple partition support. The Power 730 supports disk-only I/O drawers (FC 5886, FC 5887, and FC EDR1), and also two 12X attached I/O drawers (FC 5802 and FC 5877), providing extensive capability to expand the overall server.

The following I/O drawers are supported on the Power 710 and Power 730 servers:

- 12X I/O drawer PCIe, SFF disk (FC 5802) Power 730 only
- 12X I/O drawer PCIe, no disk (FC 5877) Power 730 only
- EXP30S holds 1.8-inch SSDs (FC EDR1)
- EXP24S holds 2.5-inch SAS disk or SSD (FC 5887)
- EXP12S holds 3.5-inch SAS disk or SSD (FC 5886)

  FC 5886 is no longer available to order with the Power 710 and Power 730 but is supported if migrated from another server.

Tips:

- A single FC 5886 or FC 5887 drawer can be cabled to the system enclosure external SAS port when an FC EJ0F DASD backplane is installed in the Power 710 and Power 730.
- A 3 Gbps YI cable (FC 3687) is used to connect the drawer to the system enclosure external SAS port.
- FC EDR1, FC 5887, and FC 5886 drawers are not available with the 4-core processor (FC EPCE) on the Power 710.

1.6.1 12X I/O drawer PCIe expansion units

The 12X I/O drawer PCIe, SFF disk (FC 5802) and 12X I/O drawer PCIe, no disk (FC 5877) expansion units are 19-inch, rack-mountable, I/O expansion drawers that are designed to be attached to the system using 12x double date rate (DDR) cables. FC 5802 and FC 5877 are packaged as a 4U form factor. The expansion units can accommodate 10 Gen3 blind swap cassettes. These cassettes can be installed and removed without removing the drawer from the rack.

The FC 5802 I/O drawer has the following attributes:

- 18 SAS hot-swap SFF disk bays
- 10 PCIe based blind swap I/O adapter slots
- Redundant hot-swappable power and cooling units

The FC 5877 drawer is the same as FC 5802 except that it does not support any disk bays.

A maximum of two FC 5802 or FC 5877 drawers can be placed on the same 12X loop. The FC 5877 I/O drawer can be on the same loop as the FC 5802 I/O drawer. An FC 5877 drawer cannot be upgraded to an FC 5802 drawer.

Power 710 support: The Power 710 does not support connections to the FC 5802 or FC 5877 I/O drawer.

Figure 1-6 shows the front view of the FC 5802 I/O drawer.

Figure 1-6 Front view of the FC 5802 I/O drawer

1.6.2 EXP30 Ultra SSD I/O drawer

The EXP30 Ultra SSD I/O Drawer (FC EDR1) provides the Power 710 and Power 730 up to 30 solid-state drives (SSD) in only 1U of rack space without any PCIe slots. The drawer provides up to 480,000 IOPS and up to 11.6 TB of capacity for AIX or Linux clients. Plus up to 48 additional hard disk drives (HDDs) can be directly attached to the Ultra Drawer (still without using any PCIe slots), providing up to 43.2 TB of additional capacity in only 4U additional rack space for AIX clients. This ultra-dense SSD option is similar to the Ultra Drawer (FC 5888), which remains available to B-models and C-models of the Power 710, and Power 730.

Reminder: The previous EXP30 drawer (FC 5888) is not supported on the D-models of the Power 710 and Power 730 servers.

The EXP30 attaches to the Power 710 or Power 730 server with a GX++ adapter, FC EJ0H. Figure 1-7 show the EXP30 Ultra SSD I/O Drawer.

Figure 1-7 EXP30 Ultra SSD I/O Drawer

1.6.3 EXP24S SFF Gen2-bay drawer

The EXP24S SFF Gen2-bay drawer is an expansion drawer supporting up to 24 2.5-inch hot-swap SFF SAS HDDs on IBM POWER6®, IBM POWER6+™, POWER7 or, POWER7+ servers in 2U of 19-inch rack space. The EXP24S bays are controlled by SAS adapters or controllers attached to the I/O drawer by SAS X or Y cables.

The SFF bays of the EXP24S differ from the SFF bays of the POWER7+ system units or 12X PCIe I/O drawers (FC 5802). The EXP24S uses Gen2 or SFF-2 SAS drives that physically do not fit in the Gen1 or SFF-1 bays of the POWER7+ system unit or 12X PCIe I/O Drawers.

The EXP24S includes redundant AC power supplies and two power cords.

Figure 1-8 shows EXP24S SFF drawer.

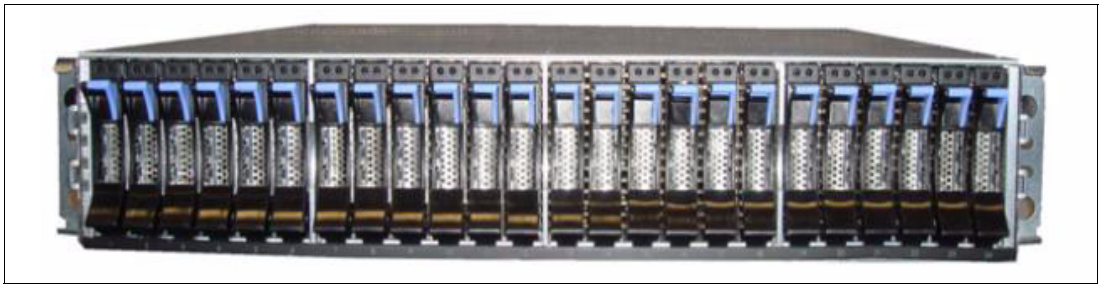


Figure 1-8 EXP24S SFF drawer

1.6.4 EXP12S SAS drawer

The EXP12S SAS drawer (FC 5886) is a 2 EIA drawer and mounts in a 19-inch rack. The drawer can hold either SAS disk drives or SSD drives. The EXP12S SAS drawer has twelve 3.5-inch SAS disk bays with redundant data paths to each bay. The SAS disk drives or SSD drives contained in the EXP12S are controlled by one or two PCIe SAS adapters that are connected to the EXP12S by using SAS cables.

The FC 5886 can also be directly attached to the SAS port on the rear of the Power 710 and Power 730, providing a low-cost disk storage solution. When used this way, the embedded SAS controllers in the system unit drive the disk drives in EXP12S. A second unit cannot be cascaded to an FC 5886 attached in this way.

The FC 5886 is no longer orderable with the Power 710 or Power 730 but is supported if migrated from another system.

Figure 1-9 shows the front view of the EXP12S SAS drawer (FC 5886).

Figure 1-9 EXP12S SAS drawer (FC 5886)

1.6.5 I/O drawers maximums

Depending on the system configuration, the maximum number of I/O drawers that is supported can vary. Table 1-9 summarizes the maximum number of I/O drawers and external disk-only I/O drawers that are supported.

Table 1-9 Maximum number of I/O drawers supported and total number of PCI slots

Server	Processor cards	Max FC 5802 and FC 5877 drawers	Max FC 5886 drawers	Max FC 5887 drawers	Max FC EDR1 drawers
Power 710	1	Not supported	8	4	Half of one
Power 730	2	2	28	14	2

Unsupported: Remember, the 4-core Power 710 does not support I/O drawers.

1.7 PowerLinux feature codes

The PowerLinux systems use the same parts as the Power 710 and Power 730 but many have a different feature codes. Table 1-10 is a cross reference between feature codes in the two types of machine.

Table 1-10 PowerLinux feature code cross reference

Feature code: Power 710 and Power 730	Feature code: PowerLinux	Description
1124	EL01	80/160 GB DAT160 SAS Tape Drive (3.5")
1751	EL35	900 GB 10K RPM SAS SFF Disk Drive (AIX, Linux)
1752	EL1R	900 GB 10K RPM SAS SFF-2 Disk Drive (AIX, Linux)
1790	EL0P	600 GB 10K RPM SAS SFF Disk Drive (AIX, Linux)
1793	EL1K	177 GB SFF-2 SSD w/ eMLC (AIX, Linux)
1818	ELQQ	Quantity 150 count of FC 1964
1866	ELQM	Quantity 150 count of FC 1917
1869	ELQN	Quantity 150 count of FC 1925
1880	EL0Z	300 GB 15K RPM SAS SFF Disk Drive (AIX, Linux)
1885	EL02	300 GB 10K RPM SFF SAS Disk Drive
1886	EL03	146 GB 15K RPM SFF SAS Disk Drive (AIX, Linux)
1887	ELQK	Quantity 150 count of FC 1793
1917	EL1M	146 GB 15K RPM SAS SFF-2 Disk Drive (AIX, Linux)
1925	EL1N	300 GB 10K RPM SAS SFF-2 Disk Drive (AIX, Linux)
1929	ELQP	Quantity 150 count of FC 1953
1953	EL1P	300 GB 15K RPM SAS SFF-2 Disk Drive (AIX, Linux)

Feature code: Power 710 and Power 730	Feature code: PowerLinux	Description
1964	EL1Q	600 GB 10K RPM SAS SFF-2 Disk Drive (AIX, Linux)
3450	EL25	SAS YO Cable 1.5 m - HD 6 Gb Adapter to Enclosure
3451	EL29	SAS YO Cable 3 m - HD 6 Gb Adapter to Enclosure
3452	EL28	SAS YO Cable 6 m - HD 6 Gb Adapter to Enclosure
3453	EL26	SAS YO Cable 10 m - HD 6 Gb Adapter to Enclosure
3454	EL1Z	SAS X Cable 3 m - HD 6 Gb 2-Adapter to Enclosure
3455	EL20	SAS X Cable 6 m - HD 6 Gb 2-Adapter to Enclosure
3456	EL1Y	SAS X Cable 10 m - HD 6 Gb 2-Adapter to Enclosure
3457	EL24	SAS YO Cable 15 m - HD 3 Gb Adapter to Enclosure
3458	EL1X	SAS X Cable 15 m - HD 3 Gb 2-Adapter to Enclosure
3661	EL22	SAS Cable (X) Adapter to SAS Enclosure, Dual Controller/Dual Path 3 m
3662	EL23	SAS Cable (X) Adapter to SAS Enclosure, Dual Controller/Dual Path 6 m
3663	EL21	SAS Cable (X) Adapter to SAS Enclosure, Dual Controller/Dual Path 15 m
3687	EL2L	SAS Cable (YI) System to SAS Enclosure, Single Controller/Dual Path 3 m
3691	EL1T	SAS Cable (YO) Adapter to SAS Enclosure, Single Controller/Dual Path 1.5 m
3692	EL1V	SAS Cable (YO) Adapter to SAS Enclosure, Single Controller/Dual Path 3 m
3693	EL1W	SAS Cable (YO) Adapter to SAS Enclosure, Single Controller/Dual Path 6 m
3694	EL1U	SAS Cable (YO) Adapter to SAS Enclosure, Single Controller/Dual Path 15 M
5260	EL11	PCIe2 LP 4-Port 1 Gb Ethernet Adapter
5265	EL0A, EL0K	Memory Riser Card
5273	EL2N	PCIe LP 8 Gb 2-Port Fibre Channel Adapter
5276	EL09	PCIe LP 4 Gb 2-Port Fibre Channel Adapter
5278	EL10	PCIe LP 2 x 4-Port SAS Adapter 3 Gb
5284	EL2P	PCIe2 LP 2-Port 10 Gb Ethernet Adapter
5802	EL36	12X I/O Drawer PCIe, SFF disk
5877	EL37	12X I/O Drawer PCIe, No Disk
5887	EL1S	EXP24S SFF Gen2-bay Drawer
5915	EL2C	SAS AA Cable 3 m - HD 6 Gb Adapter to Adapter

Feature code: Power 710 and Power 730	Feature code: PowerLinux	Description
5916	EL2D	SAS AA Cable 6 m - HD 6 Gb Adapter to Adapter
5917	EL2B	SAS AA Cable 1.5 m - HD 6 Gb Adapter to Adapter
5918	EL2A	SAS AA Cable 0.6 m - HD 6 Gb Adapter to Adapter
EC27	EL27	PCIe2 LP 2-Port 10 GbE RoCE SFP+ Adapter
EC29	EL2Z	PCIe2 LP 2-Port 10 GbE RoCE SR Adapter
EDR1	EL30	EXP30 Ultra SSD I/O Drawer
EJ0D	EL0R, EL0W	Storage Backplane: 6 SFF Drives/SATA DVD
EJ0E	EL0T, EL0X	Storage Backplane: 3 SFF Drives/SATA DVD/HH Tape
EJ0F	EL0V, EL0Y	Storage Backplane: 6 SFF Drives/SATA DVD/RAID/External SAS Port
EM08	EL15, EL1F	8 GB (2 x 4 GB) Memory DIMMs, 1066 MHz, 2 Gb DDR3 DRAM
EM4B	EL2Q, EL2S	16 GB (2 x 8 GB) Memory DIMMs, 1066 MHz, 4 Gb DDR3 DRAM
EM4C	EL2R, EL2T	32 GB (2 x 16GB) Memory DIMMs, 1066 MHz, 4 Gb DDR3 DRAM
EM4D	EL2U	64 GB (2 x 32 GB) Memory DIMMs, 1066 MHz, 4 Gb DDR3 DRAM
EN05	EL31	PCIe x8 Cable 1.5 m
EN07	EL32	PCIe x8 Cable 3 m
EN0J	EL38	PCIe2 LP 4-Port (10 Gb FCoE & 1 Gb Ethernet) SR & RJ45
EPCE	EPLP	4-core 3.6 GHz POWER7+ Processor Module
EPCG	EPLQ	6-core 4.2 GHz POWER7+ Processor Module
EPCH	EPLJ	8-core 3.6 GHz POWER7+ Processor Module
EPCJ	EPLK	8-core 4.2 GHz POWER7+ Processor Module
EPDF	EPLR	One Processor Activation for Processor Feature FC EPCF
EPDG	EPLS	One Processor Activation for Processor Feature FC EPCG
EPDH	EPLM	One Processor Activation for Processor Feature FC EPCH
EPDJ	EPLN	One Processor Activation for Processor Feature FC EPCJ
EQ0C	ELQL	Quantity of 150 FC ES0C
EQ52	ELQR	Quantity 150 of FC 1752 (900 GB SFF-2 disk)
ES02	EL34	387 GB 1.8-inch SAS SSD for AIX, Linux with eMLC
ES0C	EL1L	387 GB SFF-2 SSD for AIX, Linux with eMLC
ESA2	EL2K	PCIe2 LP RAID SAS Adapter Dual-Port 6 Gb

1.8 Build to order

You can perform a build-to-order or an *a la carte* configuration by using the IBM configurator for e-business (e-config), 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 adapter.

Preferably, begin with one of the available starting configurations, such as the IBM Edition. These solutions are available at initial system order time with a starting configuration that is ready to run as is.

1.9 IBM Edition

Each IBM Edition is available only as an initial order. If you order a Power 710 or Power 730 Express server, IBM Edition as defined next, you can qualify for half the initial configuration's processor core activations at no additional charge.

The total memory (based on the number of cores) and the quantity and size of disk, SSD, Fibre Channel adapters, or Fibre Channel over Ethernet (FCoE) adapters that are included with the server are the only features that determine whether a customer is entitled to a processor activation at no additional charge.

With an IBM Edition for a Power 710, processor activations for the processor card options are as follows:

- 3.6 GHz 4-core processor module (FC EPCE) with 2 x FC EPDE (chargeable) and 2 x FC EPEE (no-charge)
- 4.2 GHz 6-core processor module (FC EPCG) with 3 x FC EPDG (chargeable) and 3 x FC EPEG (no-charge)
- 4.2 GHz 8-core processor module (FC EPCJ) with 4 x FC EPDJ (chargeable) and 4 x FC EPEJ (no-charge)

With an IBM Edition for the Power 730, processor activations for the processor card options are as follows:

- Two 4.3 GHz 4-core processor module (FC EPCF) with 4 x FC EPDF (chargeable) and 4 x FC EPEF (no-charge)
- Two 4.2 GHz 6-core processor module (FC EPCG) with 6 x EPDG (chargeable) and 6 x FC EPEG (no-charge)
- Two 3.6 GHz 8-core processor module (FC EPCH) with 8 x FC EPDH (chargeable) and 8 x FC EPEH (no-charge)
- Two 4.2 GHz 8-core processor module (FC EPCJ) with 8 x FC EPDJ (chargeable) and 8 x FC EPEJ (no-charge)

The Power 730 requires two processor modules.

When you purchase an IBM Edition, you can purchase an AIX, IBM i, or Linux operating system license, or you can choose to purchase the system with no operating system. The AIX, IBM i, or Linux operating system is processed by means of a feature code on one of the following systems:

- AIX 6.1, or AIX 7.1
- IBM i 6.1.1 or IBM i 7.1
- SUSE Linux Enterprise Server or Red Hat Enterprise Linux

If you choose AIX 6.1 or AIX 7.1 for your primary operating system, you can also order IBM i 6.1.1 or IBM i 7.1 and SUSE Linux Enterprise Server or Red Hat Enterprise Linux. The converse is true if you choose an IBM i or Linux subscription as your primary operating system.

These sample configurations can be changed as needed and still qualify for processor entitlements at no additional charge. However, selection of total memory, HDD, SSD, Fibre Channel, or FCoE adapter quantities smaller than the totals defined as the minimums disqualifies the order as an IBM Edition, and the no-charge processor activations are then removed.

Consider the following minimum definitions for IBM Edition:

- For the Power 710, a minimum of 2 GB memory per core is needed to qualify for the IBM Edition. Various valid memory configurations can meet the minimum requirement.
- For the Power 730, a minimum of 4 GB memory per core is needed to qualify for the IBM Edition. Various valid memory configurations can meet the minimum requirement.

Additionally, a minimum of two HDDs, two SSDs, two Fibre Channel adapters, or two FCoE adapters is required. You only need to meet one of these disk, SSD, Fibre Channel, or FCoE criteria. Partial criteria cannot be combined.

- Two SAS HDDs: Any capacity drives that are located in the system unit, FC 5886 DASD drawer, or FC 5887 DASD drawer, or FC 5802 I/O drawer (Power 730 only) qualify.
- Two SAS SSDs: Any capacity drives that are located in the system unit, FC EDR1 I/O drawer, FC 5886 DASD drawer, or FC 5887 DASD drawer or FC 5802 I/O drawer (Power 730 only) qualify.
- Two SSD Modules with eMLC (FC 1995 or FC 1996): Modules that are located in the system unit with FC 2053 qualify.
- Two Fibre Channel PCI Express adapters located in the system unit or FC 5802 or FC 5877 I/O drawer (Only Power 730 supports I/O drawers).
- Two Fibre Channel over Ethernet PCI Express adapters located in the system unit or FC 5802 or FC 5877 I/O drawer (Only Power 730 supports I/O drawers).

1.10 Server and virtualization management

If you want to implement partitions, a Hardware Management Console (HMC) or the Integrated Virtualization Manager (IVM) is required to manage the Power 710 and Power 730 servers. In general, multiple IBM POWER6, POWER6+, POWER7, and POWER7+ processor-based servers can be supported by a single HMC.

Remember: If you do not use an HMC or IVM, the Power 710 and Power 730 run in full system partition mode. That means that a single partition owns all the server resources, and only one operating system can be installed.

If an HMC is used to manage the Power 710 and Power 730, the HMC must be a rack-mount CR3 or later, or deskside C05 or later.

In 2012, IBM announced a new HMC model, machine type 7042-CR7. Hardware features on the CR7 model include a second disk drive (FC 1998) for RAID 1 data mirroring, and the option of a redundant power supply. At the time of writing, the latest version of HMC code was V7R7.7.0 (SP1).

This code level also includes the new LPAR function support, which allows the HMC to manage more LPARs per processor core. A core can now be partitioned in up to 20 LPARs (0.05 of a core).

Several HMC models are supported to manage POWER7+ processor-based systems. The model 7042-CR7 is the only HMC available for ordering at the time of writing, but you can also use one of the withdrawn models listed in Table 1-11.

Table 1-11 HMC models that support POWER7+ processor technology-based servers

Type-model	Availability	Description
7310-C05	Withdrawn	IBM 7310 Model C05 Desktop Hardware Management Console
7310-C06	Withdrawn	IBM 7310 Model C06 Deskside Hardware Management Console
7042-C06	Withdrawn	IBM 7042 Model C06 Deskside Hardware Management Console
7042-C07	Withdrawn	IBM 7042 Model C07 Deskside Hardware Management Console
7042-C08	Withdrawn	IBM 7042 Model C08 Deskside Hardware Management Console
7310-CR3	Withdrawn	IBM 7310 Model CR3 Rack-Mounted Hardware Management Console
7042-CR4	Withdrawn	IBM 7042 Model CR4 Rack-Mounted Hardware Management Console
7042-CR5	Withdrawn	IBM 7042 Model CR5 Rack-Mounted Hardware Management Console
7042-CR6	Withdrawn	IBM 7042 Model CR6 Rack mounted Hardware Management Console
7042-CR7	Available	IBM 7042 Model CR7 Rack mounted Hardware Management Console

The IBM POWER7+ processor-based Power 710 and IBM Power 730 servers require HMC V7R7.7.0 Service Pack 1.

The HMC V7.7.0 (SP1) contains the following support, improvements, and abilities:

- Support for managing IBM Power 710 and Power 730
- Support for PowerVM functions such as new HMC GUI interface for VIOS installation
- Improved transition from IVM to HMC management
- Support for 802.1 Qbg on virtual Ethernet adapters
- Ability to update the user's password in Kerberos from the HMC for clients who use remote HMC

Tip: You can download or order the latest HMC code from the Fix Central website:

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

Existing HMC models 7310 can be upgraded to Licensed Machine Code Version 7 to support environments that can include IBM POWER5, IBM POWER5+, POWER6, POWER6+, POWER7, and POWER7+ processor-based servers. Licensed Machine Code Version 6 (FC 0961) is not available for 7042 HMCs.

When IBM Systems Director is used to manage an HMC, or if the HMC manages more than 254 partitions, the HMC must have a minimum of 3 GB RAM and must be a rack-mount CR3 model or later or deskside C06 or later.

1.11 System racks

The Power 710 and Power 730 are designed to mount in the 25U 7014-S25 (FC 0555), 36U 7014-T00 (FC 0551), or the 42U 7014-T42 (FC 0553) rack. These racks are built to the 19-inch EIA standard.

Order information: A new Power 710 or Power 730 server can be ordered with the appropriate 7014 rack model. The racks are available as features of the Power 710 and Power 730 only when an additional external disk drawer for an existing system (MES order) is ordered. Use the rack feature code if IBM manufacturing must integrate the newly ordered external disk drawer in a 19-inch rack before shipping the miscellaneous equipment specification (MES) order.

If a system is to be installed in a rack or cabinet that is not IBM, ensure that the rack meets the requirements described in 1.11.10, “OEM rack” on page 36.

Responsibility: The client is responsible for ensuring that the installation of the drawer in the preferred rack or cabinet results in a configuration that is stable, serviceable, safe, and compatible with the drawer requirements for power, cooling, cable management, weight, and rail security.

1.11.1 IBM 7014 Model S25 rack

The 1.3 Meter (49-inch) Model S25 rack has the following features:

- 25 EIA units
- Weights
  - Base empty rack: 100.2 kg (221 lb.)
  - Maximum load limit: 567.5 kg (1250 lb.)

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

1.11.2 IBM 7014 Model T00 rack

The 1.8-meter (71-inch) model T00 is compatible with past and present IBM Power Systems servers. The features of the T00 rack are as follows:

- Has 36U (EIA units) of usable space.
- Has optional removable side panels.
- Has optional side-to-side mounting hardware for joining multiple racks.
- Has increased power distribution and weight capacity.
- Supports both AC and DC configurations.
- Up to four power distribution units (PDUs) can be mounted in the PDU bays (see Figure 1-11 on page 29), but others can fit inside the rack. See 1.11.7, “The AC power distribution unit and rack content” on page 28.

- For the T00 rack three door options are available:
  - Front Door for 1.8 m Rack (FC 6068)

    This feature provides an attractive black full height rack door. The door is steel, with a perforated flat front surface. The perforation pattern extends from the bottom to the top of the door to enhance ventilation and provide some visibility into the rack.

    OEM front door: This door is also available as an OEM front door (FC 6101).

  - A 1.8 m Rack Acoustic Door (FC 6248)

    This feature provides a front and rear rack door designed to reduce acoustic sound levels in a general business environment.

  - A 1.8 m Rack Trim Kit (FC 6263)

    If no front door will be used in the rack, this feature provides a decorative trim kit for the front.

- Ruggedized Rack Feature

  For enhanced rigidity and stability of the rack, the optional Ruggedized Rack Feature (FC 6080) provides additional hardware that reinforces the rack and anchors it to the floor. This hardware is designed primarily for use in locations where earthquakes are a concern. The feature includes a large steel brace or truss that bolts into the rear of the rack.

  It is hinged on the left side so it can swing out of the way for easy access to the rack drawers when necessary. The Ruggedized Rack Feature also includes hardware for bolting the rack to a concrete floor or a similar surface, and bolt-in steel filler panels for any unoccupied spaces in the rack.

- Weights are as follows:
  - T00 base empty rack: 244 kg (535 lb)
  - T00 full rack: 816 kg (1795 lb)
  - Maximum weight of drawers is 572 kg (1260 lb)
  - Maximum weight of drawers in a zone 4 earthquake environment is 490 kg (1080 lb). This number equates to 13.6 kg (30 lb) per EIA.

Important: If additional weight is added to the top of the rack, for example adding FC 6117, the 490 kg (1080 lb) must be reduced by the weight of the addition. As an example, FC 6117 weighs approximately 45 kg (100 lb) so the new maximum weight of drawers that the rack can support in a zone 4 earthquake environment is 445 kg (980 lb). In the zone 4 earthquake environment, the rack must be configured starting with the heavier drawers at the bottom of the rack.

1.11.3 IBM 7014 Model T42 rack

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

- The T42 rack has 42U (EIA units) of usable space (6U of additional space).
- The model T42 supports AC power only.
- Weights are as follows:
  - T42 base empty rack: 261 kg (575 lb)
  - T42 full rack: 930 kg (2045 lb)

The available door options for T42 rack are shown in Figure 1-10.

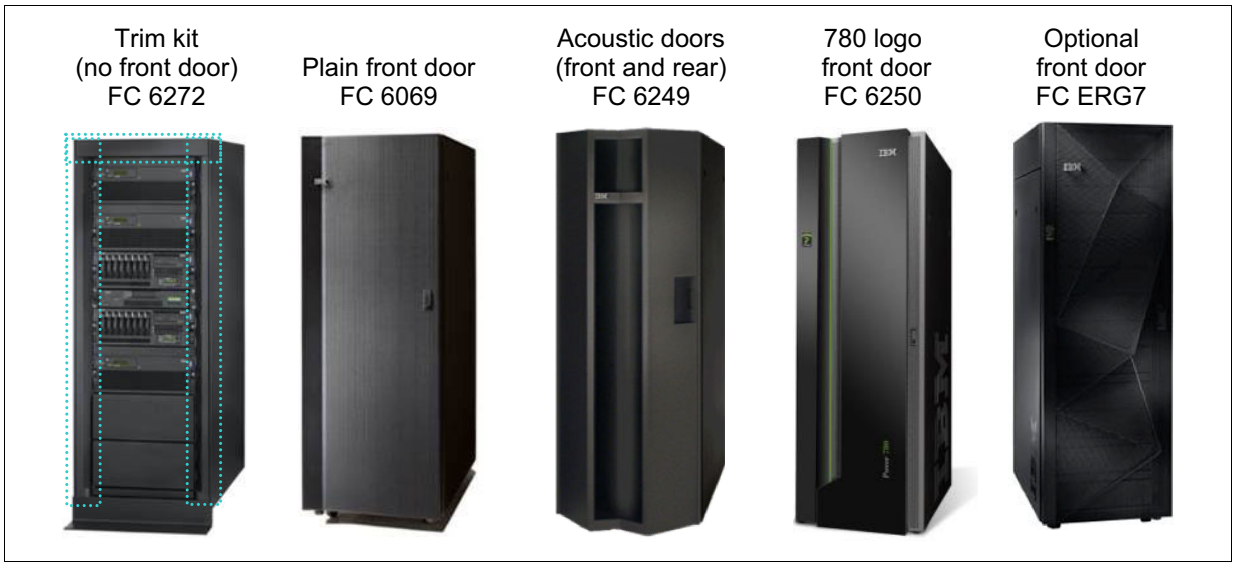


Figure 1-10 Door options for the T42 rack

- The 2.0 m Rack Trim Kit (FC 6272) is used, if no front door is used in the rack.
- The Front Door for a 2.0 m Rack (FC 6069) is made of steel, with a perforated flat front surface. The perforation pattern extends from the bottom to the top of the door to enhance ventilation and provide some visibility into the rack. This door is non acoustic and has a depth of about 25 mm (1 in).

  OEM front door: This door is also available as an OEM front door (FC 6084).

- The 2.0 m Rack Acoustic Door (FC 6249) consists of a front and rear door to reduce noise by approximately 6 dB(A). It has a depth of approximately 191 mm (7.5 in).
- The High-End Appearance Front Door (FC 6250) provides a front rack door with a field-installed Power 780 logo indicating that the rack contains a Power 780 system. The door is not acoustic and has a depth of about 90 mm (3.5 in).

  High end: For the High-End Appearance Front Door (FC 6250), use the High-End Appearance Side Covers (FC 6238) to make the rack appear as though it is a high-end server (but in a 19-inch rack format instead of a 24-inch rack).

- The FC ERG7 provides an attractive black full height rack door. The door is steel, with a perforated flat front surface. The perforation pattern extends from the bottom to the top of the door to enhance ventilation and provide some visibility into the rack. The non-acoustic door has a depth of about 134 mm (5.3 in).

Rear Door Heat Exchanger

To lead away more heat, a special door named the Rear Door Heat Exchanger (FC 6858) is available. This door replaces the standard rear door on the rack. Copper tubes that are attached to the rear door circulate chilled water, provided by the customer. The chilled water removes heat from the exhaust air being blown through the servers and attachments mounted in the rack. With industry standard quick couplings, the water lines in the door attach to the customer-supplied secondary water loop.

For details about planning for the installation of the IBM Rear Door Heat Exchanger, see the following website:

http://pic.dhe.ibm.com/infocenter/powersys/v3r1m5/index.jsp?topic=/iphad_p5/iphadexchangeroverview.html

1.11.4 Feature code 0555 rack

The 1.3 Meter Rack (FC 0555) is a 25 EIA unit rack. The rack that is delivered as FC 0555 is the same rack that is delivered when you order the 7014-S25 rack. The included features might vary. The FC 0555 is supported, but no longer orderable.

1.11.5 Feature code 0551 rack

The 1.8 Meter Rack (FC 0551) is a 36 EIA unit rack. The rack that is delivered as FC 0551 is the same rack that is delivered when you order the 7014-T00 rack. The included features might vary. Certain features that are delivered as part of the 7014-T00 must be ordered separately with the FC 0551.

1.11.6 Feature code 0553 rack

The 2.0 Meter Rack (FC 0553) is a 42 EIA unit rack. The rack that is delivered as FC 0553 is the same rack that is delivered when you order the 7014-T42 or B42 rack. The included features might vary. Certain features that are delivered as part of the 7014-T42 or B42 must be ordered separately with the FC 0553.

1.11.7 The AC power distribution unit and rack content

For rack models T00 and T42, 12-outlet PDUs are available. These include the AC power distribution units FC 9188 and FC 7188 and the AC Intelligent PDU+ FC 5889 and FC 7109.

The Intelligent PDU+ (FC 5889 and FC 7109) is identical to FC 9188 and FC 7188 PDUs but are equipped with one Ethernet port, one console serial port, and one RS232 serial port for power monitoring.

The PDUs have 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.

Four PDUs can be mounted vertically in the back of the T00 and T42 racks. Figure 1-11 shows 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, the best approach is to use fillers in the EIA units that are occupied by these PDUs to facilitate proper air-flow and ventilation in the rack.

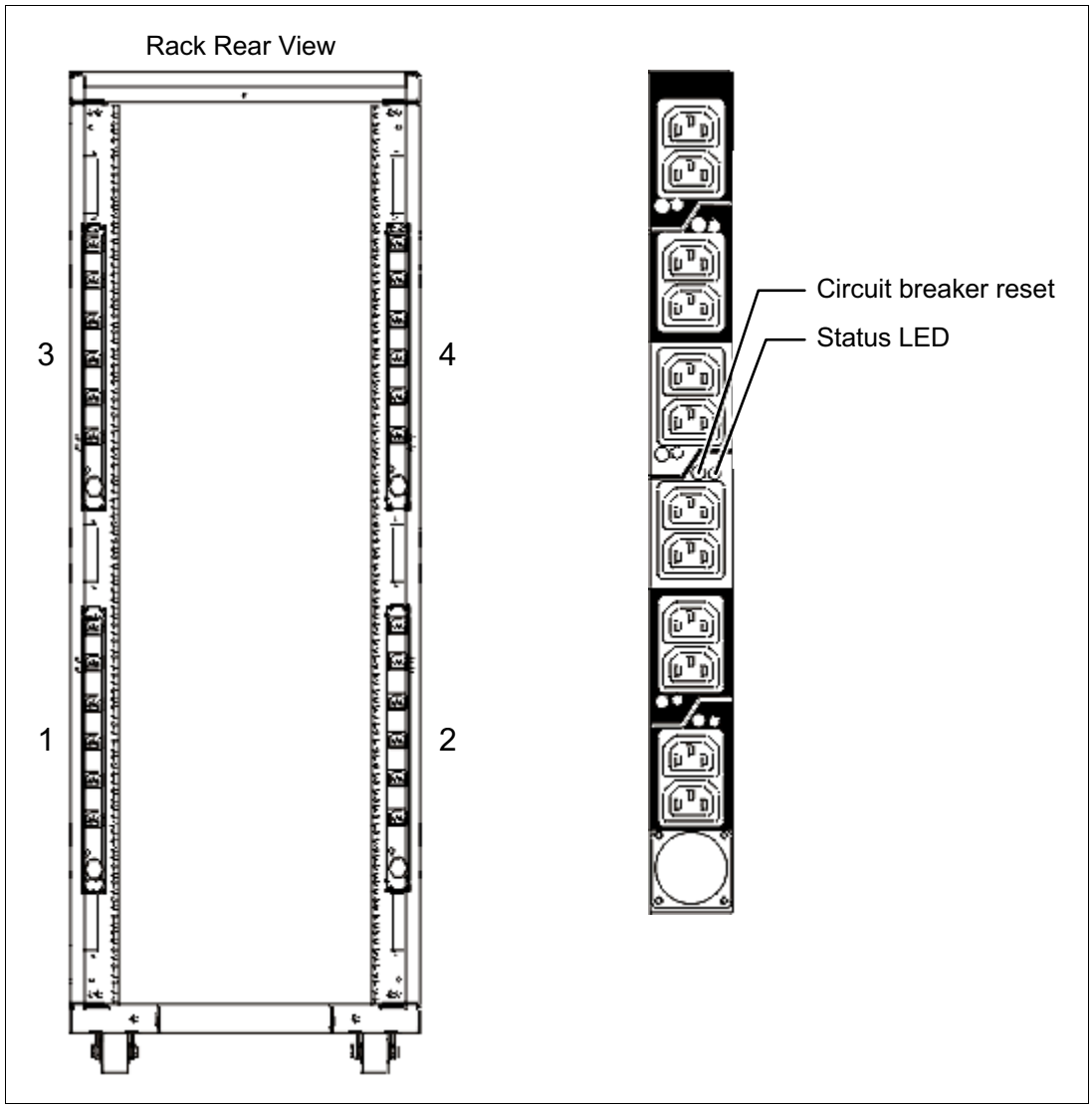


Figure 1-11 PDU placement and PDU view

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 various 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.

Table 1-12 shows the available wall power cord options for the PDU and iPDU features, which must be ordered separately.

Table 1-12 Wall power cord options for the PDU and iPDU features

Feature code	Wall plug	Rated voltage (Vac)	Phase	Rated amperage	Geography
6653	IEC 309, 3P+N+G, 16A	230	3	16 Amps	Internationally available
6489	IEC309 3P+N+G, 32A	230	3	24 Amps	EMEA
6654	NEMA L6-30	200-208, 240	1	24 Amps	US, Canada, LA, Japan
6655	RS 3750DP (watertight)	200-208, 240	1	24 Amps	US, Canada, LA, Japan
6656	IEC 309, P+N+G, 32A	230	1	24 Amps	EMEA
6657	PDL	230-240	1	24 Amps	Australia, New Zealand
6658	Korean plug	220	1	24 Amps	North and South Korea
6492	IEC 309, 2P+G, 60A	200-208, 240	1	48 Amps	US, Canada, LA, Japan
6491	IEC 309, P+N+G, 63A	230	1	48 Amps	EMEA

Notes: Ensure that the appropriate power cord feature is configured to support the power being supplied. Based on the power cord that is used, the PDU can supply from 4.8 kVA to 19.2 kVA. The power of all the drawers plugged into the PDU must not exceed the power cord limitation.

The Universal PDUs are compatible with previous models.

To better enable electrical redundancy, each server has two power supplies that must be connected to separate PDUs, which are not included in the base order.

Redundant power supplies: The second power supply for the Power 710 server is optional and not included in the base order.

For maximum availability, a highly desirable approach is to connect power cords from the same system to two separate PDUs in the rack, and to connect each PDU to independent power sources.

For detailed power requirements and power cord details, see the Planning for power section in the IBM Power Systems Hardware Information Center website:

http://pic.dhe.ibm.com/infocenter/powersys/v3r1m5/topic/p7had/p7hadrpower.htm

1.11.8 Rack-mounting rules

Consider the following primary rules when you mount the system into a rack:

- The system is designed to be placed at any location in the rack. For rack stability, start filling a rack from the bottom.
- Any remaining space in the rack can be used to install other systems or peripherals, if 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, be sure to follow the rack manufacturer's safety instructions regarding rack stability.

1.11.9 Useful rack additions

This section highlights several solutions for IBM Power Systems rack-based systems.

IBM System Storage 7214 Tape and DVD Enclosure

The IBM System Storage® 7214 Tape and DVD Enclosure (Model 1U2) is designed to mount in one EIA unit of a standard IBM Power Systems 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.

Table 1-13 shows the supported tape or DVD drives for IBM Power servers in the 7214-1U2:

Table 1-13 Supported feature codes for 7214-1U2

Feature code	Description	Status
1400	DAT72 36 GB Tape Drive	Available
1401	DAT160 80 GB Tape Drive	Available
1402	DAT320 160 GB SAS Tape Drive	Withdrawn
1420	DVD-RAM SAS Optical Drive	Available
1421	DVD-ROM Optical Drive	Withdrawn
1423	DVD-ROM Optical Drive	Available
1404	LTO Ultrium 4 Half-High 800 GB Tape Drive	Available

Support: The IBM System Storage 7214-1U2 Tape and DVD Enclosure is no longer orderable, although the drawer is supported to be attached to a Power 710 or Power 730 server.

IBM System Storage 7216 Multi-Media Enclosure

The IBM System Storage 7216 Multi-Media Enclosure (Model 1U2) is designed to attach to the Power 710 and the Power 730 through a USB port on the server or through a PCIe SAS adapter. The 7216 has two bays to accommodate external tape, removable disk drive, or DVD-RAM drive options.

Table 1-14 shows the supported tape, RDX, or DVD drives for IBM Power servers in the 7216-1U2.

Table 1-14 Supported feature codes for 7216-1U2

Feature code	Description	Status
5619	DAT160 80 GB SAS Tape Drive	Available
EU16	DAT160 80 GB USB Tape Drive	Available
1402	DAT320 160 GB SAS Tape Drive	Withdrawn
5673	DAT320 160 GB USB Tape Drive	Withdrawn
1420	DVD-RAM SAS Optical Drive	Withdrawn
8247	LTO Ultrium 5 Half-High 1.5 TB SAS Tape Drive	Withdrawn
1103	RDX Removable Disk Drive Docking Station	Withdrawn

Support: The IBM System Storage 7216-1U2 Multi-Media Enclosure is no longer orderable, although the drawer is supported to be attached to a Power 710 or Power 730 server.

To attach a 7216 Multi-Media Enclosure to the Power 710 and Power 730, consider the following cabling procedures:

- Attachment by an SAS adapter

  A PCIe Dual-X4 SAS adapter (FC 5901) or a PCIe LP 2-x4-Port SAS Adapter 3 Gb (FC 5278) must be installed in the Power 710 and Power 730 server to attach to a 7216 Model 1U2 Multi-Media Storage Enclosure. Attaching a 7216 to a Power 710 and Power 730 through the integrated SAS adapter is not supported.

  For each SAS tape drive and DVD-RAM drive feature installed in the 7216, the appropriate external SAS cable will be included.

  An optional Quad External SAS cable is available by specifying (FC 5544) with each 7216 order. The Quad External Cable allows up to four 7216 SAS tape or DVD-RAM features to attach to a single System SAS adapter.

  Up to two 7216 storage enclosure SAS features can be attached per PCIe Dual-X4 SAS adapter (FC 5901) or the PCIe LP 2-x4-Port SAS Adapter 3 Gb (FC 5278).

- Attachment by a USB adapter

  The Removable RDX HDD Docking Station features on 7216 support only the USB cable that is provided as part of the feature code. Additional USB hubs, add-on USB cables, or USB cable extenders are not supported.

  For each RDX Docking Station feature installed in the 7216, the appropriate external USB cable will be included. The 7216 RDX Docking Station feature can be connected to the external, integrated USB ports on the Power 710 and Power 730 or to the USB ports on 4-Port USB PCI Express Adapter (FC 2728).

  The 7216 DAT320 USB tape drive or RDX Docking Station features can be connected to the external integrated USB ports on the Power 710 and Power 730.

The two drive slots of the 7216 enclosure can hold the following drive combinations:

- One tape drive (DAT160 SAS or LTO Ultrium 5 Half-High SAS) with second bay empty
- Two tape drives (DAT160 SAS or LTO Ultrium 5 Half-High SAS) in any combination
- One tape drive (DAT160 SAS or LTO Ultrium 5 Half-High SAS) and one DVD-RAM SAS drive sled with one or two DVD-RAM SAS drives
- Up to four DVD-RAM drives
- One tape drive (DAT160 SAS or LTO Ultrium 5 Half-High SAS) in one bay, and one RDX removable HDD docking station in the other drive bay
- One RDX removable HDD docking station and one DVD-RAM SAS drive sled with one or two DVD-RAM SAS drives in the bay on the right
- Two RDX removable HDD docking stations

Figure 1-12 shows the 7216 Multi-Media Enclosure.

Figure 1-12 7216 Multi-Media Enclosure

In general, the 7216-1U2 is supported by the AIX, IBM i, and Linux operating systems. IBM i, from Version 7.1, now fully supports the internal 5.25 inch RDX SATA removable HDD docking station, including boot support (no VIOS support). This support provides a fast, robust, high-performance alternative to tape backup and restore devices.

IBM System Storage 7226 Model 1U3 Multi-Media Enclosure

IBM System Storage 7226 Model 1U3 Multi-Media Enclosure can accommodate up to two tape drives, two RDX removable disk drive docking stations, or up to four DVD RAM drives. The 7226 offers SAS, USB, and FC electronic interface drive options.

The 7226 Storage Enclosure delivers external tape, removable disk drive, and DVD-RAM drive options that allow data transfer within similar system archival storage and retrieval technologies that are installed in existing IT facilities. The 7226 offers an expansive list of drive feature options.

Table 1-15 shows the supported options for IBM Power servers in the 7226-1U3.

Table 1-15 Supported feature codes for 7226-1U3

Feature code	Description	Status
5619	DAT160 SAS Tape Drive	Available
EU16	DAT160 USB Tape Drive	Available
1420	DVD-RAM SAS Optical Drive	Available
5762	DVD-RAM USB Optical Drive	Available
8248	LTO Ultrium 5 Half High Fibre Drive	Available
8247	LTO Ultrium 5 Half High SAS Drive	Available
8348	LTO Ultrium 6 Half High Fibre Drive	Available
EU11	LTO Ultrium 6 Half High SAS Drive	Available
1103	RDX 2.0 Removable Disk Docking Station	Withdrawn
EU03	RDX 3.0 Removable Disk Docking Station	Available

Option descriptions are as follows:

- DAT160 80 GB Tape Drives: With SAS or USB interface options and a data transfer rate of up to 24 MBps, the DAT160 drive is read/write compatible with DAT160, DAT72, and DDS4 data cartridges.
- LTO Ultrium 5 Half-High 1.5 TB SAS and FC Tape Drive: With a data transfer rate up to 280 MBps, the LTO Ultrium 5 drive is read/write compatible with LTO Ultrium 5 and LTO Ultrium 4 data cartridges, and read-only compatible with Ultrium 3 data cartridges. By using data compression, an LTO-5 cartridge is capable of storing up to 3 TB of data.
- LTO Ultrium 6 Half-High 2.5 TB SAS and FC Tape Drive: With a data transfer rate up to 160 MBps, the LTO Ultrium 6 drive is read/write compatible with LTO Ultrium 5 and LTO Ultrium 4 data cartridges. By using data compression, an LTO-6 cartridge is capable of storing up to 6.25 TB of data.
- DVD-RAM: The 9.4 GB SAS Slim Optical Drive with an SAS and USB interface option is compatible with most standard DVD disks.
- RDX removable disk drives: The RDX USB docking station is compatible with most RDX removable disk drive cartridges when used in the same operating system. The 7226 offers the following RDX removable drive capacity options:
  - 320 GB (FC EU08)
  - 500 GB (FC 1107)
  - 1.0 TB (FC EU01)
  - 1.5 TB (FC EU15)

Removable RDX drives are in a rugged cartridge that inserts in an RDX removable (USB) disk docking station (FC 1103 or FC EU03). RDX drives are compatible with docking stations, installed internally in IBM POWER6, POWER6+, POWER7, and POWER7+ servers.

Media used in the 7226 DAT160 SAS and USB tape drive features are compatible with DAT160 tape drives installed internally in IBM POWER6, POWER6+, POWER7, and POWER7+ servers, and in IBM BladeCenter® systems.

Media used in LTO Ultrium 5 Half-High 1.5 TB tape drives are compatible with Half High LTO5 tape drives installed in the IBM TS2250 and TS2350 external tape drives, IBM LTO5 tape libraries, and half-high LTO5 tape drives installed internally in IBM POWER6, POWER6+, POWER7, and POWER7+ servers.

Figure 1-12 shows the 7226 Multi-Media Enclosure.

Figure 1-13 FC 7226 Multi-Media Enclosure

The 7226 enclosure offers customer-replaceable unit (CRU) maintenance service to help make installation or replacement of new drives efficient. Other 7226 components are also designed for CRU maintenance.

The IBM System Storage 7226 Multi-Media Enclosure is compatible with most IBM POWER6, POWER6+, POWER7, and POWER7+ systems, and also with the IBM BladeCenter models (PS700, PS701, PS702, PS703, and PS704) that offer current level AIX, IBM i, and Linux operating systems.

Unsupported: The IBM i operating system does not support 7226 USB devices.

For a complete list of host software versions and release levels that support the 7226, see the following System Storage Interoperation Center (SSIC) website:

`http://www.ibm.com/systems/support/storage/config/ssic/index.jsp`

Flat panel display options

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

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

1.11.10 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. For detailed information, see the IBM Power Systems Hardware Information Center at the following website:

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

The website mentions the following key points:

- 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). Figure 1-14 is a top view showing the specification dimensions.

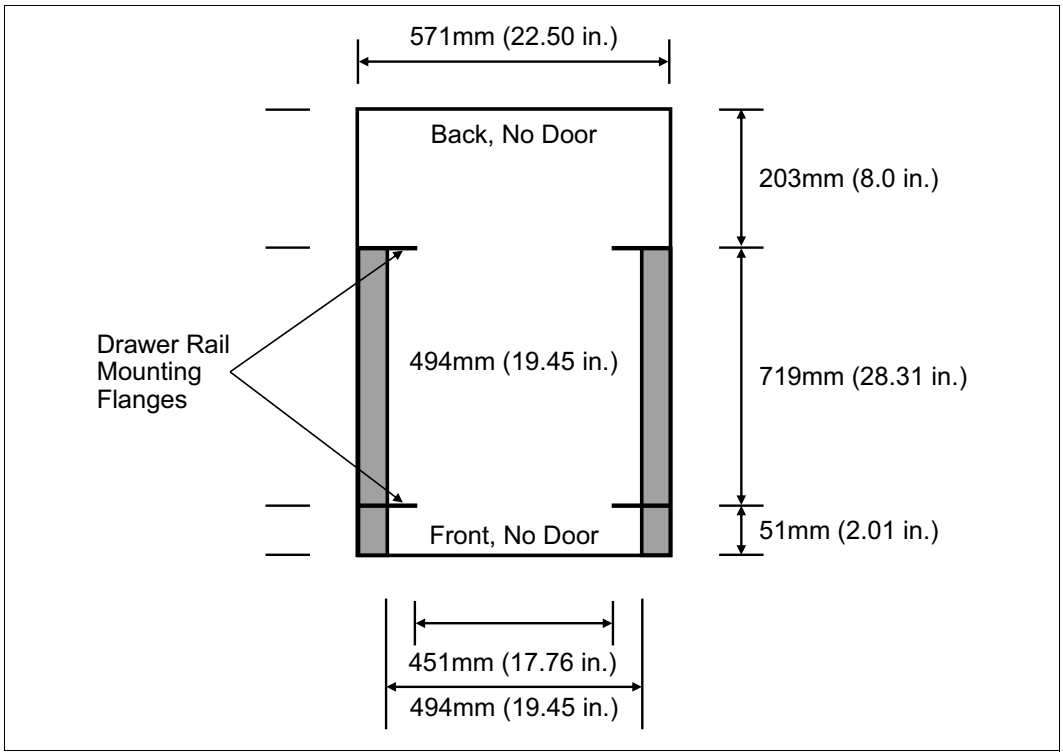


Figure 1-14 Top view of rack specification dimensions (not specific to IBM)

- 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. Figure 1-15 shows the top front specification dimensions.

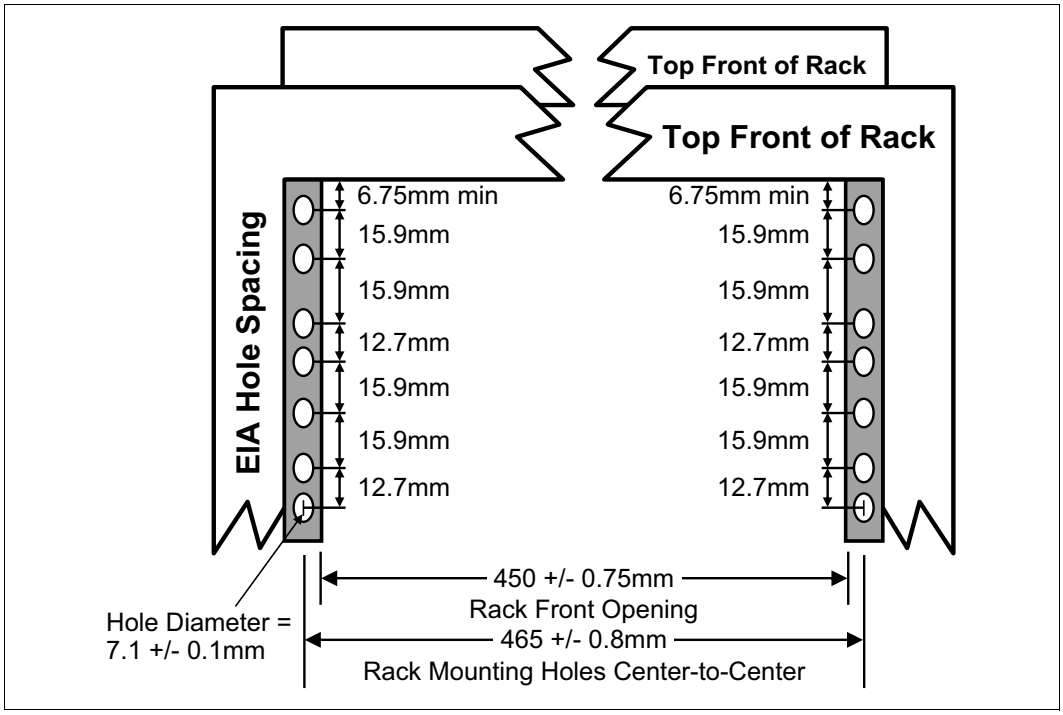


Figure 1-15 Rack specification dimensions, top front view

2

Architecture and technical overview

This chapter discusses the overall system architecture for the IBM Power 710 and Power 730, represented by Figure 2-1 on page 40 and Figure 2-2 on page 41. The bandwidths that are provided throughout the section 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.

Always do the performance sizing at the application workload environment level and evaluate performance using real-world performance measurements and production workloads.

© Copyright IBM Corp. 2013. All rights reserved.

Figure 2-1 shows the logical system diagram for the Power 710.

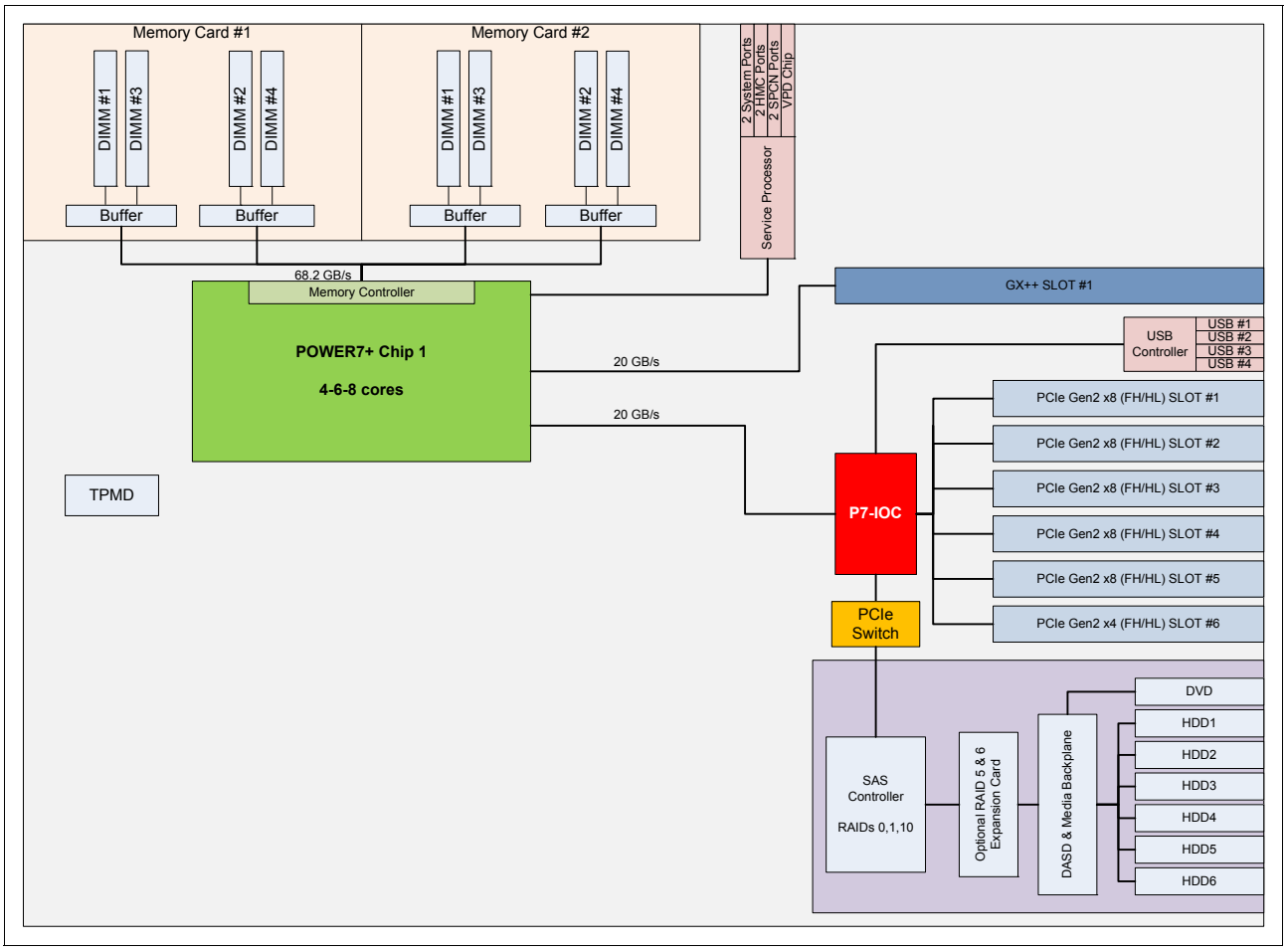


Figure 2-1 IBM Power 710 logical system diagram

Figure 2-2 shows the logical system diagram for the Power 730.

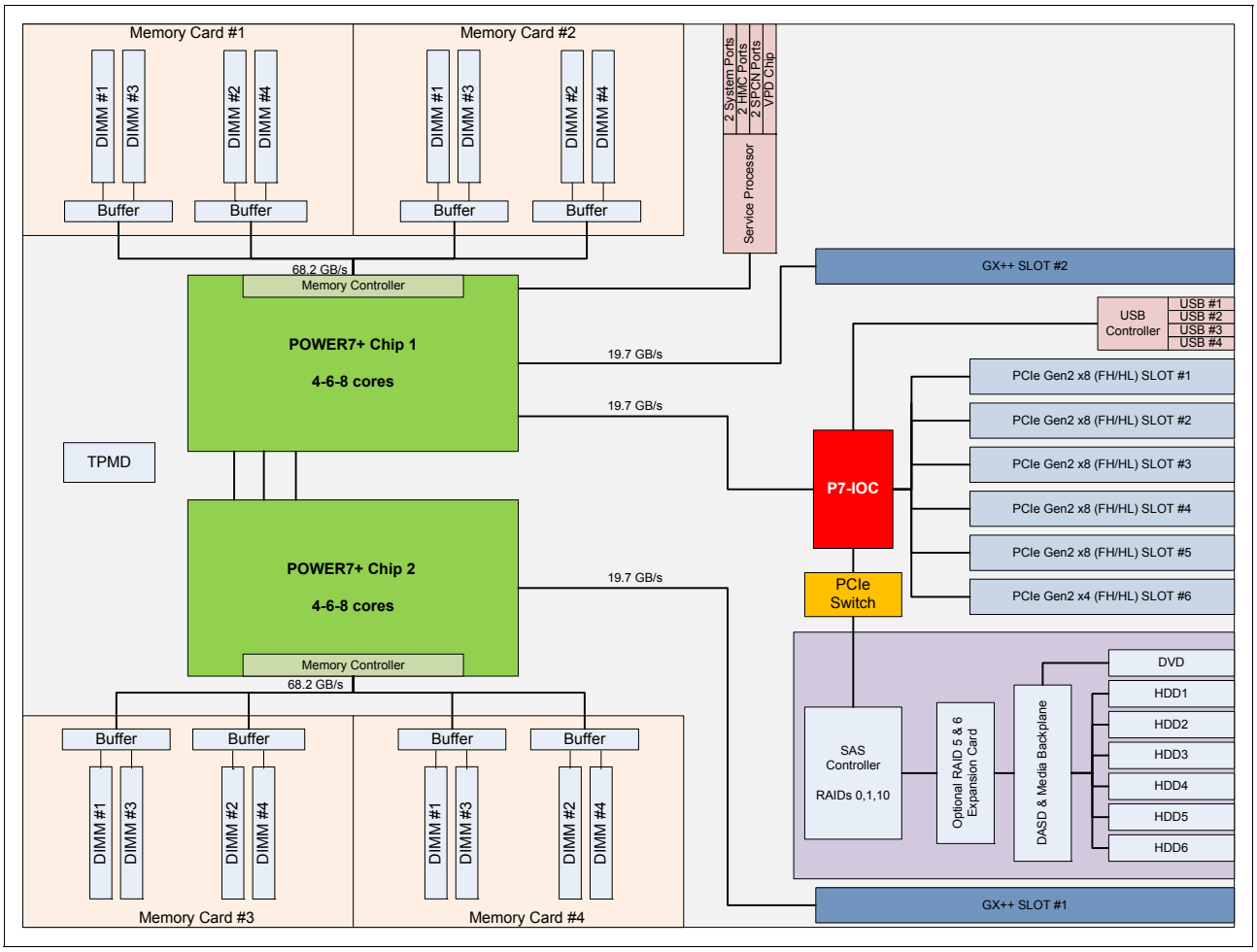


Figure 2-2 IBM Power 730 logical system diagram

2.1 The IBM POWER7+ processor

The IBM POWER7+ processor represents a leap forward in technology achievement and associated computing capability. The multi-core architecture of the POWER7+ processor is matched with innovation across a wide range of related technologies to deliver leading throughput, efficiency, scalability, and reliability, availability, and serviceability (RAS).

Although the processor is an important component in delivering outstanding servers, many elements and facilities must be balanced on a server to deliver maximum throughput. As with previous generations of systems based on IBM POWER® processors, the design philosophy for POWER7+ processor-based systems is one of system-wide balance in which the POWER7+ processor plays an important role.

IBM uses innovative technologies to achieve required levels of throughput and bandwidth. Areas of innovation for the POWER7+ processor and POWER7+ processor-based systems include (but are not limited to) the following items:

- On-chip L3 cache implemented in embedded dynamic random access memory (eDRAM)
- Cache hierarchy and component innovation
- Advances in memory subsystem
- Advances in off-chip signaling
- Advances in I/O card throughput and latency
- Advances in RAS features such as power-on reset and L3 cache dynamic column repair

The superscalar POWER7+ processor design also provides a variety of other capabilities:

- Binary compatibility with the prior generation of POWER processors
- Support for PowerVM virtualization capabilities, including PowerVM Live Partition Mobility to and from POWER6, POWER6+, and POWER7 processor-based systems

Figure 2-3 on page 43 shows the POWER7+ processor die layout, with the major areas identified:

- Processor cores
- L2 cache
- L3 cache and chip interconnection
- Simultaneous multiprocessing links
- Memory controllers.
- I/O links

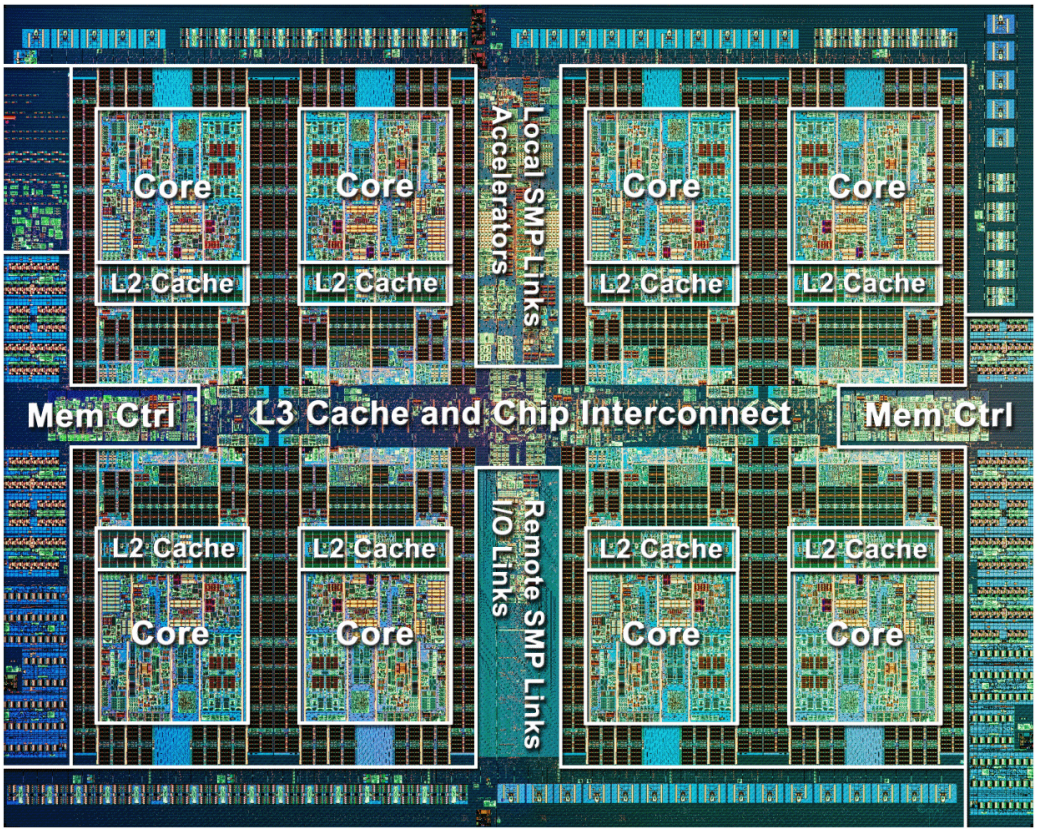


Figure 2-3 POWER7+ processor die with key areas indicated

2.1.1 POWER7+ processor overview

The POWER7+ processor chip is fabricated with IBM 32 nm Silicon-On-Insulator (SOI) technology using copper interconnects, and implements an on-chip L3 cache using eDRAM.

The POWER7+ processor chip is 567 mm^2 and has 2.1 billion components (transistors). Up to eight processor cores are on the chip, each with 12 execution units, 256 KB of L2 cache per core, and up to 80 MB of shared on-chip L3 cache per chip.

For memory access, the POWER7+ processor includes a double data rate 3 (DDR3) memory controller with four memory channels.

Table 2-1 summarizes the technology characteristics of the POWER7+ processor.

Table 2-1 Summary of POWER7+ processor technology

Technology	POWER7+ processor
Die size	567 mm^2
Fabrication technology	► 32 nm lithography ► Copper interconnect ► Silicon-on-Insulator ► eDRAM
Processor cores	3, 4, 6, or 8
Maximum execution threads core/chip	4/32
Maximum L2 cache core/chip	256 KB/2 MB
Maximum On-chip L3 cache core/chip	10 MB/80 MB
DDR3 memory controllers	1
SMP design-point	32 sockets with IBM POWER7+ processors
Compatibility	With prior generation of POWER processor

2.1.2 POWER7+ processor core

Each POWER7+ processor core implements aggressive out-of-order (OoO) instruction execution to drive high efficiency in the use of available execution paths. The POWER7+ processor has an Instruction Sequence Unit that is capable of dispatching up to six instructions per cycle to a set of queues. Up to eight instructions per cycle can be issued to the instruction execution units. The POWER7+ processor has a set of 12 execution units:

- Two fixed point units
- Two load store units
- Four double precision floating point units
- One vector unit
- One branch unit
- One condition register unit
- One decimal floating point unit

The following caches are tightly coupled to each POWER7+ processor core:

- Instruction cache: 32 KB
- Data cache: 32 KB
- L2 cache: 256 KB, implemented in fast SRAM

2.1.3 Simultaneous multithreading

POWER7+ processors support SMT1, SMT2, and SMT4 modes to enable up to four instruction threads to execute simultaneously in each POWER7+ processor core. The processor supports the following instruction thread execution modes:

- SMT1: Single instruction execution thread per core
- SMT2: Two instruction execution threads per core
- SMT4: Four instruction execution threads per core

SMT4 mode enables the POWER7+ processor to maximize the throughput of the processor core by offering an increase in processor-core efficiency. SMT4 mode is the latest step in an evolution of multithreading technologies introduced by IBM.

Figure 2-4 shows the evolution of simultaneous multithreading in the industry.

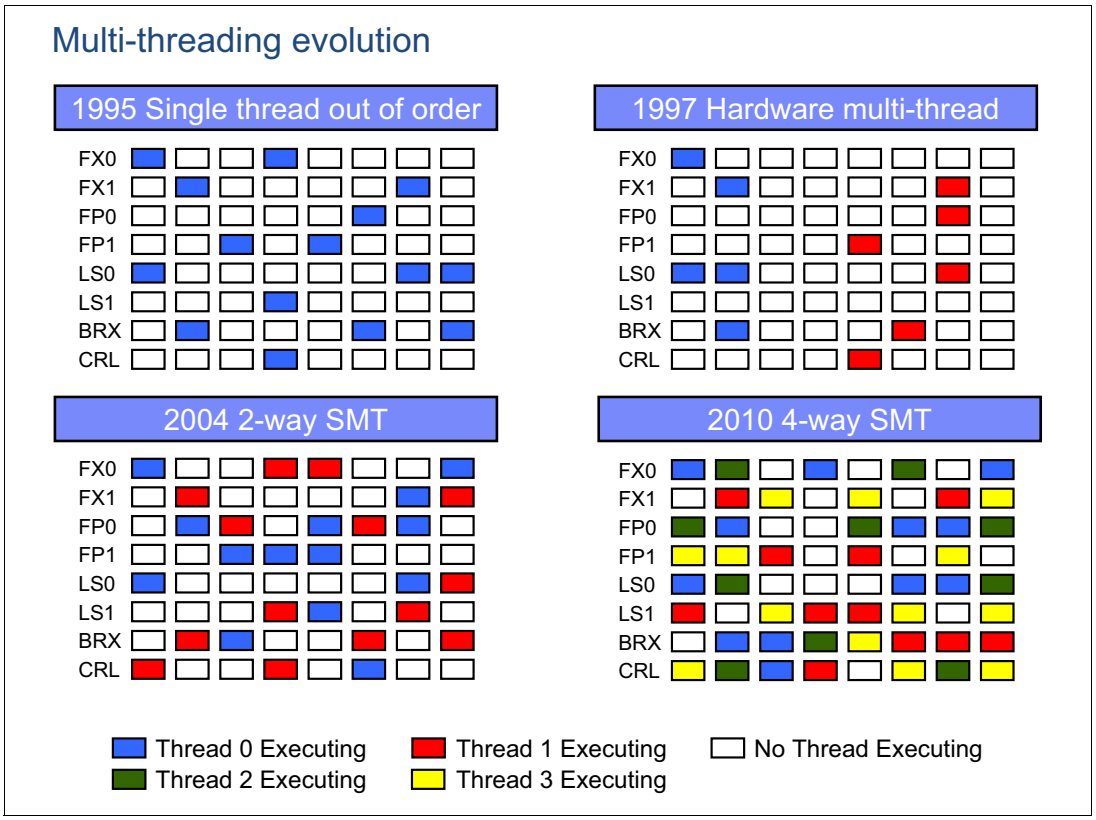


Figure 2-4 Evolution of simultaneous multithreading

The various SMT modes offered by the POWER7+ processor allow flexibility, enabling users to select the threading technology that meets an aggregation of objectives such as performance, throughput, energy use, and workload enablement.

Intelligent Threads

The POWER7+ processor features Intelligent Threads that can vary based on the workload demand. The system either automatically selects (or the system administrator can manually select) whether a workload benefits from dedicating as much capability as possible to a single thread of work, or if the workload benefits more from having capability spread across two or four threads of work. With more threads, the POWER7+ processor can deliver more total capacity as more tasks are accomplished in parallel.

With fewer threads, those workloads that need fast individual tasks can get the performance that they need for maximum benefit.

2.1.4 Memory access

Each POWER7+ processor chip has one memory controller that uses four memory channels. Each memory channel operates at 1066 MHz connects to two DIMMs.

In the Power 710 server, each channel can address up to 64 GB. Thus the Power 710 is capable of addressing up to 256 GB of total memory.

In the Power 730 server, each channel can address up to 64 GB. Thus the Power 730 is capable of addressing up to 512 GB of total memory.

Figure 2-5 gives a simple overview of the POWER7+ processor memory access structure in the Power 710 and Power 730.

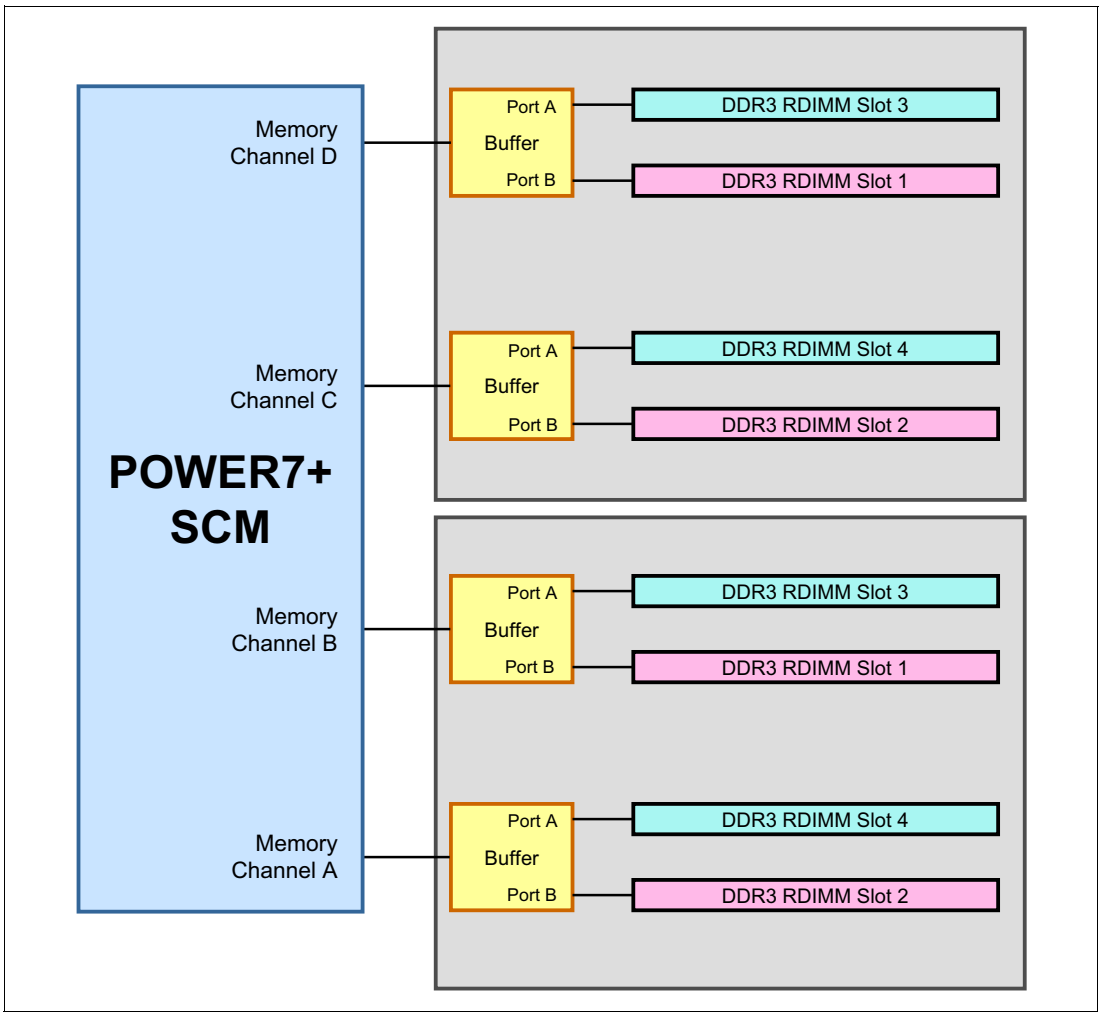


Figure 2-5 Overview of POWER7+ memory access structure

2.1.5 On-chip L3 cache innovation and Intelligent Cache

A breakthrough in material engineering and microprocessor fabrication enabled IBM to implement the L3 cache in eDRAM and place it on the POWER7+ processor die. L3 cache is critical to a balanced design, as is the ability to provide good signaling between the L3 cache and other elements of the hierarchy, such as the L2 cache or SMP interconnect.

The on-chip L3 cache is organized into separate areas with differing latency characteristics. Each processor core is associated with a fast local region of L3 cache (FLR-L3) but also has access to other L3 cache regions as shared L3 cache. Additionally, each core can negotiate to use the FLR-L3 cache associated with another core, depending on reference patterns. Data can also be cloned to be stored in more than one core's FLR-L3 cache, again depending on reference patterns. This Intelligent Cache management enables the POWER7+ processor to optimize the access to L3 cache lines and minimize overall cache latencies.

Figure 2-6 shows fast local L3 cache region for each core on the POWER7+ processor die.

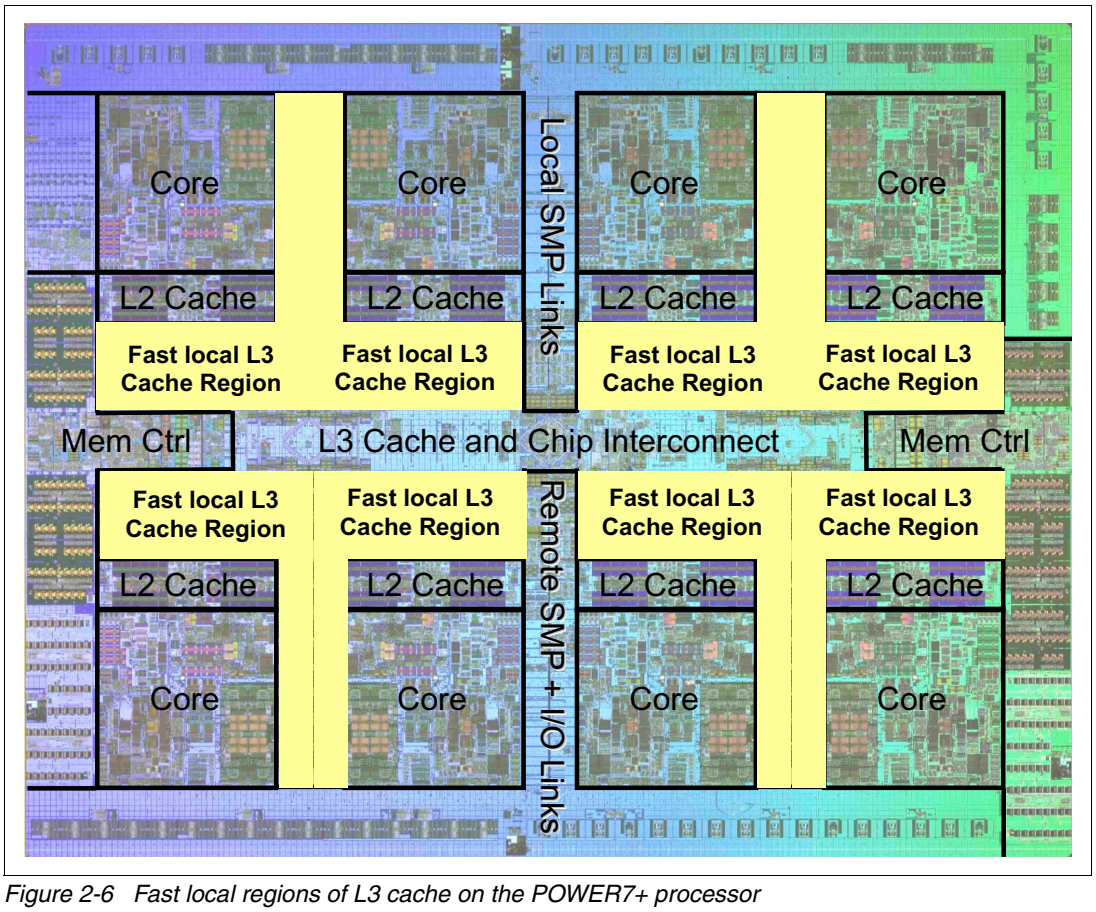


Figure 2-6 Fast local regions of L3 cache on the POWER7+ processor

The innovation of using eDRAM on the POWER7+ processor die is significant for several reasons:

- Latency improvement

  A six-to-one latency improvement occurs by moving the L3 cache on-chip compared to L3 accesses on an external (on-ceramic) ASIC.

- Bandwidth improvement

  A 2x bandwidth improvement occurs with on-chip interconnect. Frequency and bus sizes are increased to and from each core.

- No off-chip driver or receivers

  Removing drivers or receivers from the L3 access path lowers interface requirements, conserves energy, and lowers latency.

- Small physical footprint

  The performance of eDRAM when implemented on-chip is similar to conventional SRAM but requires far less physical space. IBM on-chip eDRAM uses only a third of the components that conventional SRAM uses, which has a minimum of six transistors to implement a 1-bit memory cell.

- Low energy consumption

  The on-chip eDRAM uses only 20% of the standby power of SRAM.

2.1.6 POWER7+ processor and Intelligent Energy

Energy consumption is an important area of focus for the design of the POWER7+ processor, which includes Intelligent Energy features that help to dynamically optimize energy usage and performance so that the best possible balance is maintained. Intelligent Energy features, such as EnergyScale, work with IBM Systems Director Active Energy Manager™ to dynamically optimize processor speed based on thermal conditions and system utilization.

2.1.7 Comparison of the POWER7+, POWER7, and POWER6 processors

Table 2-2 shows comparable characteristics between the generations of POWER7+, POWER7, and POWER6 processors.

Table 2-2 Comparison of technology for the POWER7+ processor and the prior generation

Characteristics	POWER7+	POWER7	POWER6
Technology	32 nm	45 nm	65 nm
Die size	567 mm^2	567 mm^2	341 mm^2
Maximum cores	8	8	2
Maximum SMT threads per core	4 threads	4 threads	2 threads
Maximum frequency	4.4 GHz	4.25 GHz	5.0 GHz
L2 Cache	256 KB per core	256 KB per core	4 MB per core

Characteristics	POWER7+	POWER7	POWER6
L3 Cache	10 MB of FLR-L3 cache per core with each core having access to the full 80 MB of L3 cache, on-chip eDRAM	4 MB or 8 MB of FLR-L3 cache per core with each core having access to the full 32 MB of L3 cache, on-chip eDRAM	32 MB off-chip eDRAM ASIC
Memory support	DDR3	DDR3	DDR2
I/O bus	Two GX++	Two GX++	One GX++
Enhanced cache mode (TurboCore)	No	Yes[a]	No

a. Only supported on the Power 795.

2.2 POWER7+ processor modules

The Power 710 and Power 730 server chassis house POWER7+ processor single chip modules (SCMs). Each SCM can access four DDR3 memory DIMM slots.

The Power 710 server houses one processor module, offering 4-core 3.6 GHz, 6-core 4.2 GHz, or 8-core 4.2 GHz configurations.

The Power 730 server houses two processor modules, offering 8-core 4.3 GHz, 12-core 4.2 GHz, and 16-core 3.6 Ghz and 4.2 GHz configurations.

All installed processors must be activated, unless they are factory deconfigured by using FC 2319.

Note: All POWER7+ processors in the system must be the same frequency and have the same number of processor cores. POWER7+ processor types cannot be mixed within a system.

2.2.1 Modules and cards

Figure 2-7 shows a Power 730 server highlighting the POWER7+ processor modules and the memory riser cards.

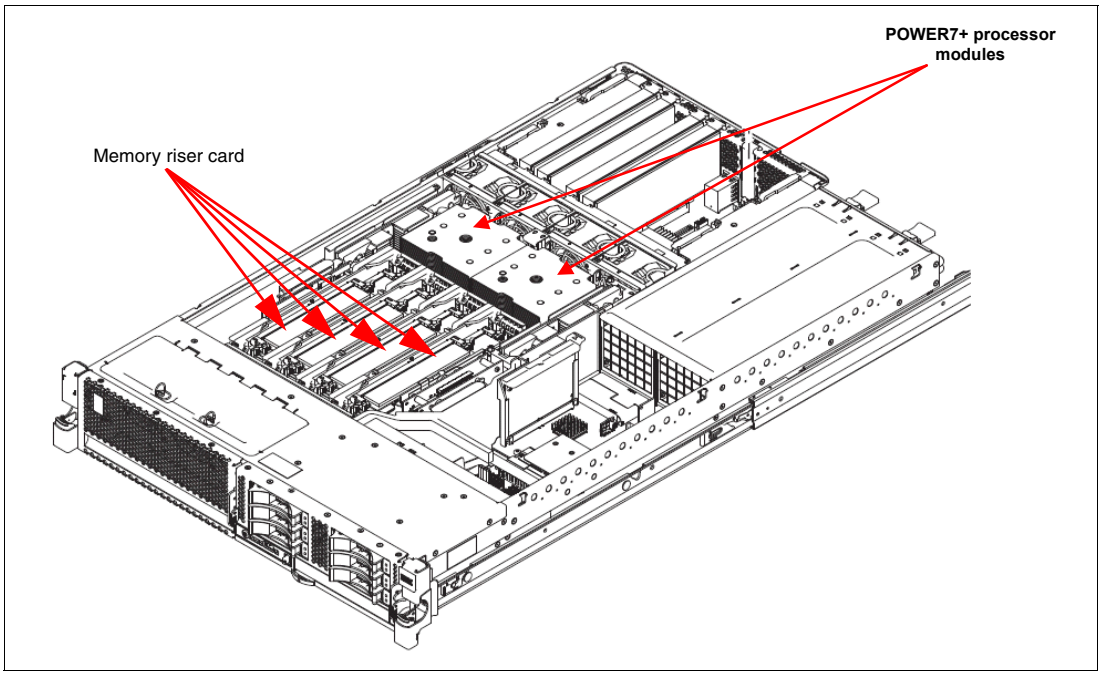


Figure 2-7 Power 730 with two POWER7+ processor modules and four memory riser cards

2.2.2 Power 710 and Power 730 systems

Power 710 and Power 730 systems support POWER7+ processors with various core-counts. Table 2-3 summarizes the POWER7+ processor options for the Power 710 system.

Table 2-3 Summary of POWER7+ processor options for the Power 710 system

Feature code	Cores per POWER7+ processor	Frequency (GHz)	Processor activation	Min/Max[a] cores per system	Min/Max[a] processor module
EPCE	4	3.6	The 4-core 3.6 GHz requires that four processor activation codes are ordered, available as 4 x FC EPDE or 2 x FC EPDE and 2 x FC EPEE.	4/4	1/1
EPCG	6	4.2	The 6-core 4.2 GHz requires that six processor activation codes be ordered, available as 6 x FC EPDG or 3 x FC EPDG and 3 x FC EPEG.	6/6	1/1
EPCJ	8	4.2	The 8-core 4.2 GHz requires that eight processor activation codes be ordered, available as 8 x FC EPDJ or 4 x FC EPDJ and 4 x FC EPEJ.	8/8	1/1

a. Minimum and maximum

Table 2-4 summarizes the POWER7+ processor options for the Power 730 system.

Table 2-4 Summary of POWER7+ processor options for the Power 730 system

Feature	Cores per POWER7 processor	Frequency (GHz)	Processor activation	Min/Max cores per system	Min/Max processor module
EPCF	4	4.3	The 4-core 4.3 GHz requires that four processor activation codes are ordered, available as 4 x FC EPDF or 2 x FC EPDF and 2 x FC EPEF	8/8	2/2
EPCG	6	4.2	The 6-core 4.2 GHz requires that six processor activation codes are ordered, available as 6 x FC EPDG or 3 x FC EPDG and 3 x FC EPEG.	12/12	2/2
EPCH	8	3.6	The 8-core 3.6 GHz requires that six processor activation codes are ordered, available as 8 x FC EPDH or 4 x FC EPDH and 4 x FC EPEH.	16/16	2/2
EPCJ	8	4.2	The 8-core 4.2 GHz requires that eight processor activation codes are ordered, available as 8 x FC EPDJ or 4 x FC EPDJ and 4 x FC EPEJ.	16/16	2/2

2.3 Memory subsystem

The Power 710 is a one-socket system that supports a single POWER7+ processor module. The server supports a maximum of eight DDR3 DIMM slots, with four DIMM slots included in the base configuration and four DIMM slots available with an optional memory riser card. The supported memory features (two memory DIMMs per feature) are 8 GB, 16 GB, 32 GB, and 64 GB, running at speeds of 1066 MHz. A system with the optionally installed memory riser card has a maximum memory of 256 GB.

The Power 730 is a two-socket system that supports up to two POWER7+ processor modules. The server supports a maximum of 16 DDR3 DIMM slots, with four DIMM slots included in the base configuration, and 12 DIMM slots available with three optional memory riser cards. The supported memory features (two memory DIMMs per feature) are 8 GB, 16 GB, 32 GB, and 64 GB, running at speeds of 1066 MHz. A system with three optionally installed memory riser cards has a maximum memory of 512 GB.

These servers support an optional feature called Active Memory Expansion (FC 4795) that allows the effective maximum memory capacity to be much larger than the true physical memory. This feature executes innovative compression or decompression of memory content by using processor cycles to provide memory expansion up to 125%, depending on the workload type and its memory utilization. A server with a maximum of 256 GB can effectively be expanded over 512 GB. This approach can enhance virtualization and server consolidation by allowing a partition to do significantly more work with the same physical amount of memory or a server to run more partitions and do more work with the same physical amount of memory.

2.3.1 Registered DIMM

Industry standard DDR3 Registered DIMM (RDIMM) technology is used to increase reliability, speed, and density of memory subsystems by putting a register between the DIMM modules and the memory controller. This register is also referred to as a buffer.

2.3.2 Memory placement rules

The following memory options are orderable:

- 8 GB (2 x 4 GB) Memory DIMMs, 1066 MHz (FC EM08)
- 16 GB (2 x 8 GB) Memory DIMMs, 1066 MHz (FC EM4B, CCIN 31FA)
- 32 GB (2 x 16 GB) Memory DIMMs, 1066 MHz (FC EM4C)
- 64 GB (2 x 32 GB) Memory DIMMs, 1066 MHz (FC EM4D)

A minimum of 8 GB memory is required for a Power 710 system or a Power 730 system.

The supported maximum memory is as follows:

- Power 710: 256 GB (four 32 GB DIMMs on each of two memory cards)
- Power 730: 512 GB (four 32 GB DIMMs on each of four memory cards)

Figure 2-8 shows the physical memory DIMM topology.

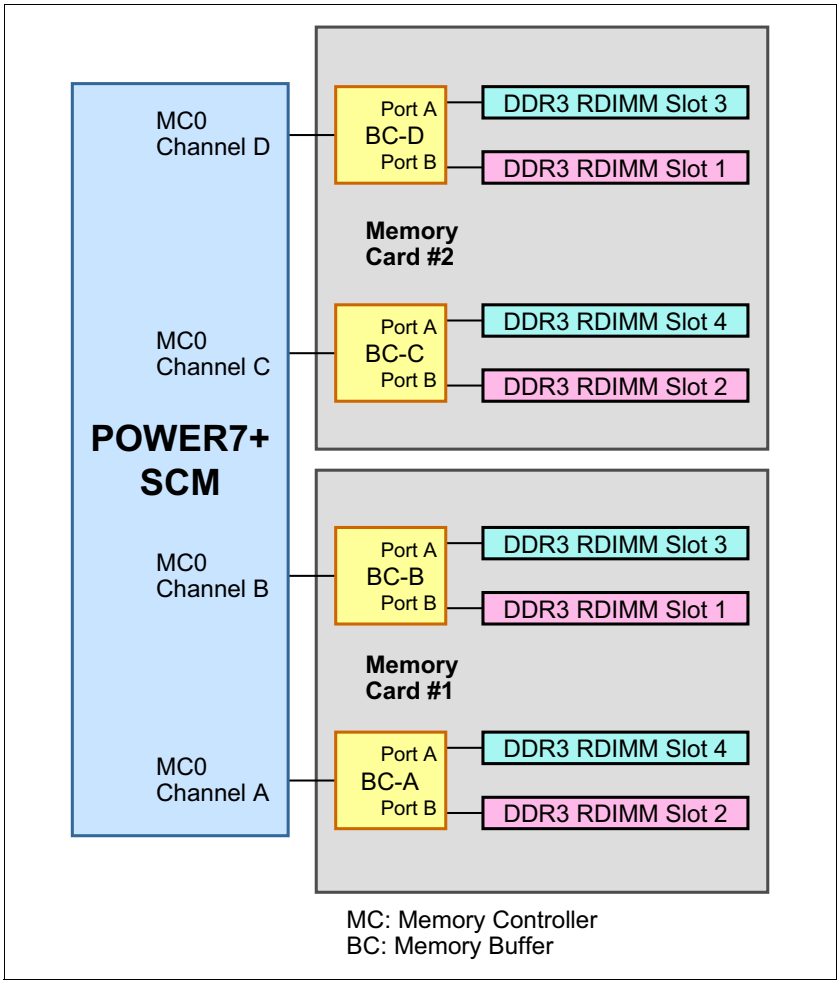


Figure 2-8 Memory DIMM topology for the Power 710 and Power 730

The memory-placement rules are as follows:

- The base machine contains one memory riser card with four DIMM sockets. Memory features occupy two memory DIMM sockets.
- The Power 710 offers one additional memory riser card feature (1 x FC 5265, CCIN 2BE3) with an additional four DIMM sockets. Maximum system memory is 128 GB without feature FC 5265 and 256 GB with one feature FC 5265.
- The Power 730 offers three optional memory riser card features (3 x FC 5265, CCIN 2BE3) with an additional four DIMM sockets per feature. Maximum system memory is 128 GB without feature FC 5265 and 256 GB with three feature FC 5265.
- A system can be ordered with a single memory feature FC EM08, FC EM4B, FC EM4C, or FC EM4D. The second memory feature, ordered on the same memory riser card, does not have to match the first memory feature. Memory features can be mixed on either memory riser card.
- A minimum of one memory feature must be plugged into each memory riser card. Empty memory riser cards are not allowed.
- There is a performance benefit when all DIMMs on a memory riser card are of the same capacity.

In general, the best approach is to install memory evenly across all memory riser cards in the system. Balancing memory across the installed memory riser cards allows memory access in a consistent manner and typically results in the best possible performance for your configuration. However, balancing memory fairly evenly across multiple memory riser cards, compared to balancing memory exactly evenly, typically has a small performance difference.

Account for any plans for future memory upgrades when you decide which memory feature size to use at the time of the initial system order.

Figure 2-9 through Figure 2-13 on page 55 show the memory DIMMs plugging rules for Power 710 and Power 730.

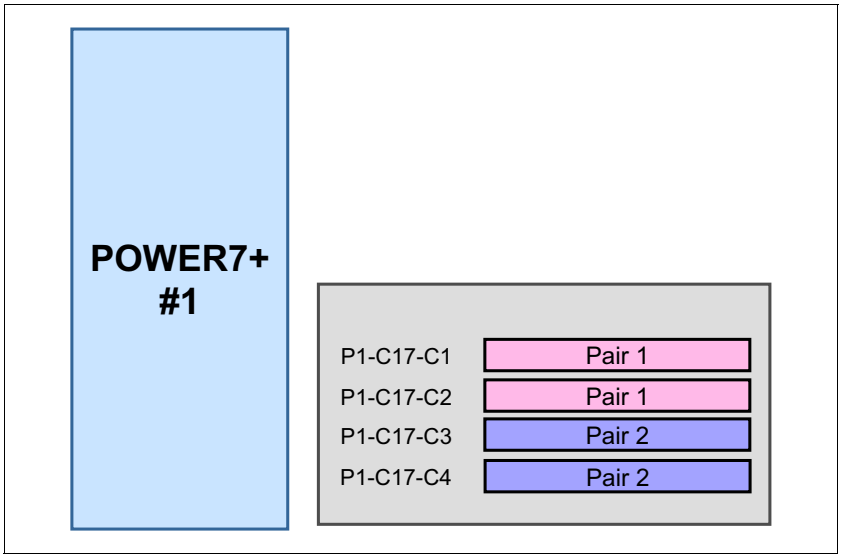


Figure 2-9 Memory DIMM installation sequence for one processor with one riser cards

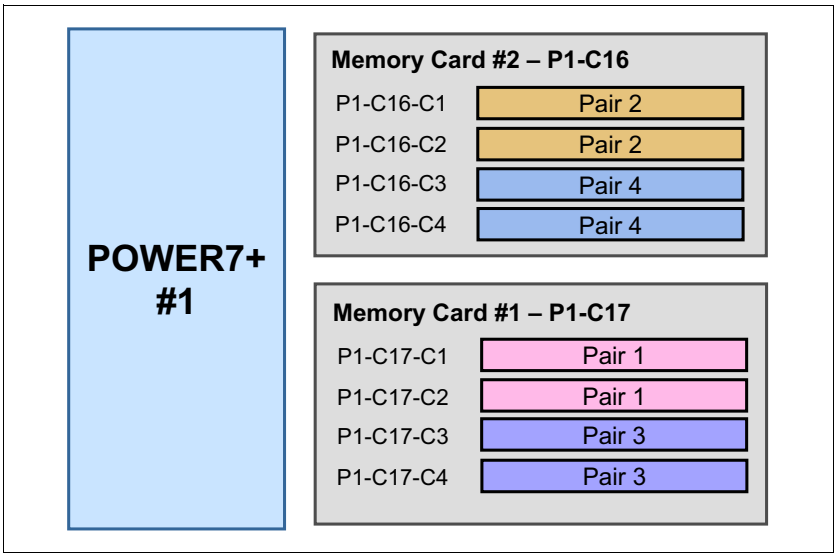


Figure 2-10 Memory DIMM installation sequence for one processor with two riser cards

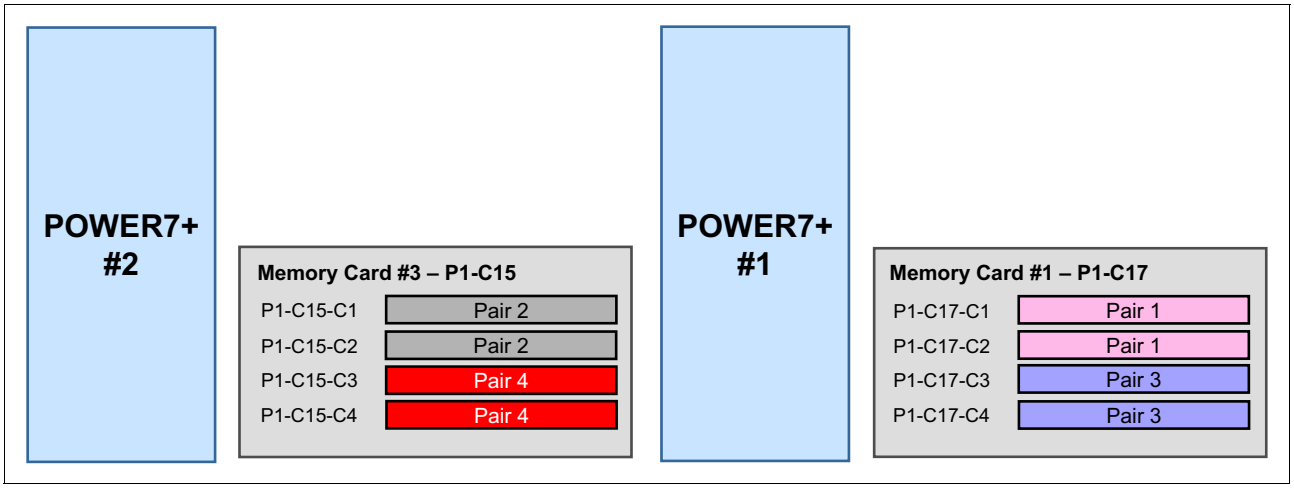


Figure 2-11 Memory DIMM installation sequence for two processor with two riser cards (Power 730 only)

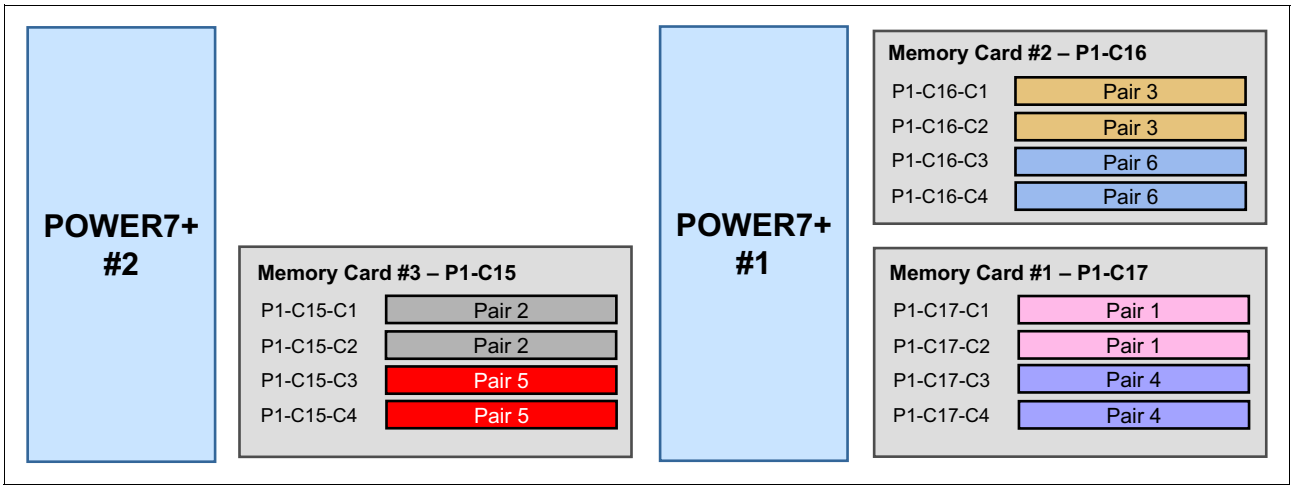


Figure 2-12 Memory DIMM installation sequence for two processor with three riser cards (Power 730 only)

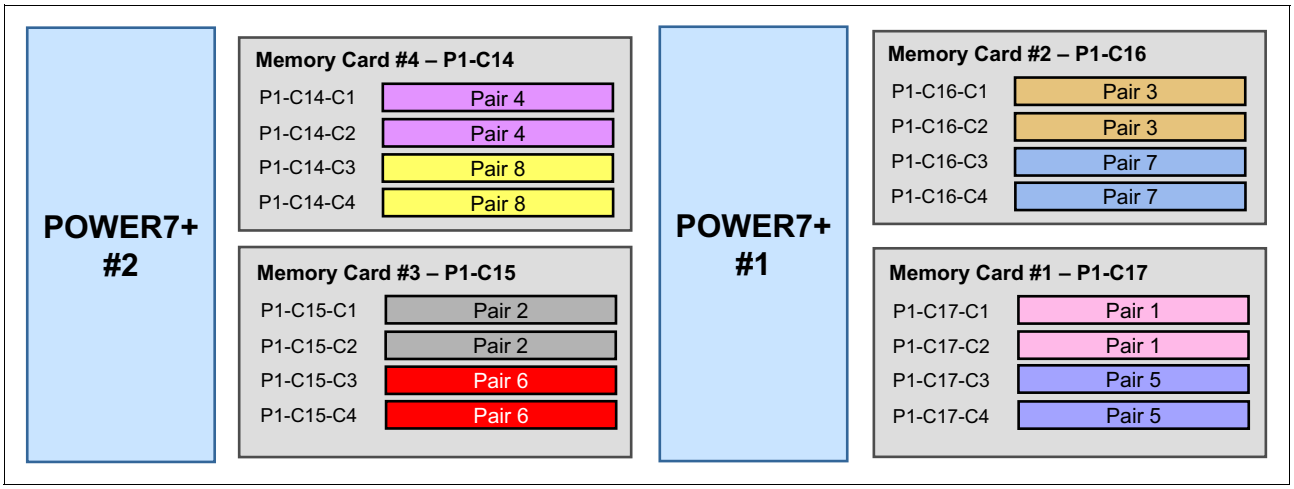


Figure 2-13 Memory DIMM installation sequence for two processor with four riser cards (Power 730 only)

Figure 2-14 shows the DIMM slot positions on the memory riser cards.

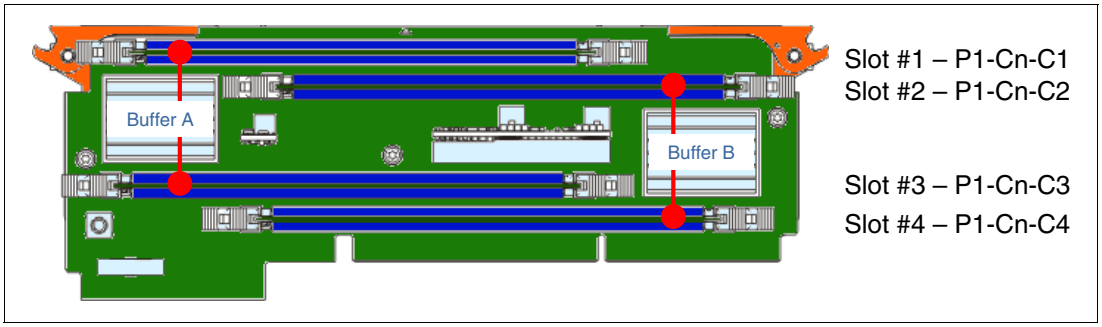


Figure 2-14 Memory riser card for the Power 710 and Power 730 systems

2.3.3 Memory bandwidth

The POWER7+ processor has exceptional cache, memory, and interconnect bandwidths. Table 2-5 shows the maximum bandwidth estimates for the Power 710 and Power 730 systems.

Table 2-5 Power 710 and Power 730 processor and memory bandwidth estimates

Memory	Power 710	Power 730
	4.284 GHz processor card	**4.312 GHz processor card**
L1 (data) cache	205.632 GBps	206.976 GBps
L2 cache	205.632 GBps	206.976 GBps
L3 cache	137.088 GBps	137.984 GBps
System memory	68.224 GBps	68.224 GBps (single socket) 136.448 GBps (dual sockets)

The bandwidth figures for the caches are calculated as follows:

- L1 cache: In one clock cycle, two 16-byte load operations and one 16-byte store operation can be accomplished. By using a 4.312 GHz processor card, the formula is as follows:

  `(2 * 16 B + 1 * 16 B) * 4.312 GHz = 206.976 GBps`

- L2 cache: In one clock cycle, one 32-byte load operation and one 16-byte store operation can be accomplished. By using a 4.312 GHz processor card the formula is as follows:

  `(1 * 32 B + 1 * 16 B) * 4.312 GHz = 206.976 GBps`

- L3 cache: One 32-byte load operation and one 32-byte store operation can be accomplished at half-clock speed. By using a 4.312 GHz processor card the formula is as follows:

  `(1 * 32 B + 1 * 32 B) * (4.312 GHz / 2) = 137.984 GBps`

- Memory: The Power 710 and Power 730 system use one memory controller of the POWER7+ processor. The memory controller is connected to a buffer chip using four ports with 8 bytes. Each buffer chip connects to two DIMMs running at 1066 MHz. See Figure 2-8 on page 52 for reference. The bandwidth formula is calculated as follows:

  `1 memory controller * 4 ports * 8 bytes * 2 DIMMs * 1066 MHz = 68.224 GBps`

2.4 Capacity on Demand

Capacity on Demand is not supported on the Power 710 and Power 730 systems.

2.5 System bus

This section provides more information related to the internal buses.

The Power 710 and Power 730 systems have internal I/O connectivity through Peripheral Component Interconnect Express (PCI Express, or PCIe) slots, and also external connectivity through InfiniBand adapters.

The internal I/O subsystem on the Power 710 and Power 730 is connected to the GX bus on a POWER7+ processor in the system. This bus runs at 2.5 GHz and provides 20 GBps of I/O connectivity to the PCIe slots, integrated Ethernet adapter ports, SAS internal adapters, and USB ports.

Additionally, the POWER7+ processor chip is installed on the Power 710 and each of the processor chips on the Power 730 provide a GX++ bus, which is used to optionally connect to the GX adapters. Each bus runs at 2.5 GHz and provides 20 GBps bandwidth.

The GX++ LP 1-port PCIe2 x8 Adapter (FC EJ0H) can be installed in the GX++ slot on the Power 710 and on either or both GX++ slots of the Power 730. It is used for attaching FC EDR1 EXP30 Ultra SSD I/O Drawer to the server. When FC EJ0H is installed in GX++ slot 2, PCIe slot 6 is physically blocked and cannot be used.

The GX++ Dual-port 12x Channel Attach Adapter (FC EJ0G) is supported only on the Power 730. It is used for Remote I/O Drawer Expansion, such as for attaching FC 5802 or FC 5877 I/O drawers to the system enclosure. FC EJ0G is a double-wide adapter that requires the installation of a separate SPCN controller card (part of FC EJ0G), which is installed in GX++ slot 1. Therefore, when installing FC EJ0G in the system, GX++ slots 1 and 2 are used and PCIe slots 5 and 6 are not usable.

Remember: The GX++ slots are not hot-pluggable.

Table 2-6 lists the I/O bandwidth of Power 710 and Power 730 processor configurations.

Table 2-6 I/O bandwidth

I/O	I/O bandwidth (maximum theoretical)
GX++ Bus from the first POWER7 SCM to the I/O chip	10 GBps simplex 20 GBps duplex
GX++ Bus (slot 1)	10 GBps simplex 20 GBps duplex
GX++ Bus (slot 2 - on Power 730 only)	10 GBps simplex 20 GBps duplex
Total I/O bandwidth	Power 710: 20 GBps simplex 40 GBps duplex Power 730: 30 GBps simplex 60 GBps duplex

2.6 Internal I/O subsystem

The internal I/O subsystem resides on the system planar, which supports PCIe slots. PCIe adapters on the Power 710 and Power 730 are not hot pluggable. However, PCIe adapters on the attached I/O drawers are hot-pluggable.

All PCIe slots support Enhanced Error Handling (EEH). PCI EEH-enabled adapters respond to a special data packet generated from the affected PCIe 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. For Linux, EEH support extends to the majority of frequently used devices, although certain third-party PCI devices might not provide native EEH support.

2.6.1 Slot configuration

Table 2-7 displays the PCIe Gen2 slot configuration of Power 710 and Power 730.

Table 2-7 Slot configuration of a Power 710 and Power 730

Slot	Description	Location code	PCI Host Bridge (PHB)	Max card size
Slot 1	PCIe Gen2 x8	P1-C2	P7IOC PCIe PHB5	Low profile
Slot 2	PCIe Gen2 x8	P1-C3	P7IOC PCIe PHB4	Low profile
Slot 3	PCIe Gen2 x8	P1-C4	P7IOC PCIe PHB3	Low profile
Slot 4	PCIe Gen2 x8	P1-C5	P7IOC PCIe PHB2	Low profile
Slot 5	PCIe Gen2 x8	P1-C6	P7IOC PCIe PHB1	Low profile
Slot 6	PCIe Gen2 x4	P1-C7	P7IOC PCIe PHB0	Low profile

Remember: Slot 6 is shared with GX++ adapter slot 2. If a PCIe adapter is plugged into slot 6, then a GX++ adapter cannot be plugged into GX++ slot 2 and the other way around.

2.6.2 System ports

The system planar has two serial ports that are called system ports. When an HMC is connected to the server, the integrated system ports of the server are rendered non-functional. In this case, you must install an asynchronous adapter, which is described in Table 2-16 on page 67, for serial port usage:

- Integrated system ports are not supported under AIX or Linux when the HMC ports are connected to an HMC. Either the HMC ports or the integrated system ports can be used, but not both.
- The integrated system ports are supported for modem and asynchronous terminal connections. Any other application using serial ports requires a serial port adapter to be installed in a PCI slot. The integrated system ports do not support IBM PowerHA® configurations.
- Configuration of the two integrated system ports, including basic port settings (baud rate, and so on), modem selection, call-home and call-in policy, can be performed with the Advanced Systems Management Interface (ASMI).

Remember: The integrated console/modem port usage just described is for systems configured as a single, system-wide partition. When the system is configured with multiple partitions, the integrated console/modem ports are disabled because the TTY console and call home functions are performed with the HMC.

2.7 PCI adapters

This section covers the various types and functionality of the PCI adapters supported with the IBM Power 710 and Power 730 systems.

2.7.1 PCI express

Peripheral Component Interconnect Express (PCIe) uses a serial interface and allows for point-to-point interconnections between devices (using a directly wired interface between these connection points). A single PCIe serial link is a dual-simplex connection that uses two pairs of wires, one pair for transmit and one pair for receive, and can transmit only one bit per cycle. These two pairs of wires are called a *lane*. A PCIe link can consist 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.

Two generations of PCIe interface are supported in Power 710 and Power 730 models:

- Gen1: Capable of transmitting at the extremely high speed of 2.5 Gbps, which gives a capability of a peak bandwidth of 2 GBps simplex on an x8 interface
- Gen2: Double the speed of the Gen1 interface, which gives a capability of a Peak bandwidth of 4 GBps simplex on an x8 interface

PCIe Gen1 slots support Gen1 adapter cards and also most of the Gen2 adapters. In this case, where a Gen2 adapter is used in a Gen1 slot, the adapter will operate at PCIe Gen1 speed. PCIe Gen2 slots support both Gen1 and Gen2 adapters. In this case, where a Gen1 card is installed into a Gen2 slot, it will operate at PCIe Gen1 speed with a slight performance enhancement. When a Gen2 adapter is installed into a Gen2 slot, it will operate at the full PCIe Gen2 speed.

The Power 710 and Power 730 system enclosure is equipped with five PCIe x8 Gen2 Low Profile slots. In addition, there is a sixth PCIe x4 dedicated to a 4-port PCIe Ethernet card that comes standard with the base unit.

All adapters support Enhanced Error Handling (EEH). PCIe adapters use a different type of slot than PCI adapters. If you attempt to force an adapter into the wrong type of slot, you might damage the adapter or the slot.

Notes:

- The PCIe2 LP 4-port 1 Gb Ethernet adapter (FC 5260) is the only PCIe adapter which is allowed at the P1-C7 PCIe x4 slot in the Power 710 and Power 730 servers. Other PCIe adapters available on the Power 710 and Power 730 are not supported in the P1-C7 slot.
- If a GX++ adapter, such as the FC EJ0G or FC EJ0H is installed at the GX++ slot 2 (P1-C8), the PCIe2 LP 4-port 1 Gb Ethernet adapter (FC 5260) has to be installed in any of the available PCIe x8 Gen2 slots.
- IBM i IOPs and PCI-X adapters are not supported in the Power 710 and Power 730 systems.

2.7.2 PCIe adapter form factors

IBM POWER7 and POWER7+ processor based servers are able to support two different form factors of PCIe adapters:

- PCIe low profile (LP) cards, which are used with the Power 710 and Power 730 PCIe slots. Low profile adapters are also used in the PCIe riser card slots of Power 720 and Power 740 servers.
- PCIe full height and full high cards that are plugged into the following servers slots:
  - Power 720 and Power 740 (in base system, five PCIe half length slots are supported.)
  - Power 750
  - Power 755
  - Power 760
  - Power 770
  - Power 780
  - Power 795
  - PCIe slots of external I/O drawers, such as FC 5802 and FC 5877

Low profile PCIe adapters cards are only supported in low profile PCIe slots, and full height and full high cards are only supported in full high slots.

Figure 2-15 lists the PCIe adapter form factors.

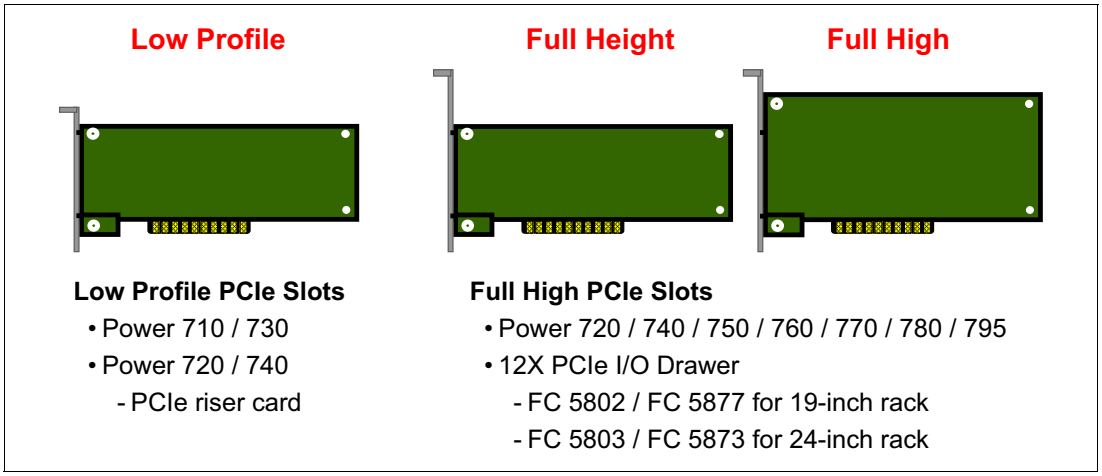


Figure 2-15 PCIe adapter form factors

Many of the full-height card features are also available in low profile format. For example, the PCIe RAID and SSD SAS Adapter 3 Gb is available as a low profile adapter or as a full height adapter, each one having a different feature code. As expected, they have equivalent functional characteristics.

Table 2-8 is a list of low-profile adapter cards and their equivalents in full height.

Table 2-8 Equivalent adapter cards

Low profile		Adapter description	Full height	
Feature code	**CCIN**		**Feature code**	**CCIN**
2053	57CD	PCIe RAID and SSD SAS Adapter 3 Gb	2054 or 2055	57CD
5269	5269	PCIe POWER GXT145 Graphics Accelerator	5748	5748
5270	2B3B	10 Gb FCoE PCIe Dual Port adapter	5708	2B3B
5271	5271	4-Port 10/100/1000 Base-TX PCI Express adapter	5717	5271
5272	5272	10 Gigabit Ethernet-CX4 PCI Express adapter	5732	5732
5273	577D	8 Gigabit PCI Express Dual Port Fibre Channel adapter	5735	577D
5274	5768	2-Port Gigabit Ethernet-SX PCI Express adapter	5768	5768
5275	5275	10 Gb ENet Fibre RNIC PCIe 8x adapter	5769	5275
5276	5774	4 Gigabit PCI Express Dual Port Fibre Channel adapter	5774	5774
5277	57D2	4-Port Sync EIA-232 PCIe adapter	5785	57D2
5278	57B3	SAS Controller PCIe 8x adapter	5901	57B3
5280	2B44	PCIe2 LP 4-Port 10 Gb Ethernet &1 Gb Ethernet SR & RJ45 adapter	5744	2B44

Low profile		Adapter description	Full height	
Feature code	CCIN		Feature code	CCIN
EN0B	577F	PCIe2 16 Gb 2-Port Fibre Channel adapter	EN0A	577F
EN0J	2B93	PCIe2 4-Port (10 Gb FCOE & 1 Gb Ethernet) SR & RJ45 adapter	EN0H	2B93

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

http://www.ibm.com/systems/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 update prerequisites to install. To do this, use the IBM prerequisite website:

https://www-912.ibm.com/e_dir/eServerPreReq.nsf

The following sections discuss the supported adapters and provide tables of orderable feature numbers. The tables indicate operating system support (AIX, IBM i, and Linux) for each of the adapters.

Support: The Power 710 and Power 730 servers support PCIe low profile adapter only. However, full height PCIe adapters can be installed in an I/O drawer (FC 5802 and FC 5877), if attached to the Power 730.

2.7.3 LAN adapters

To connect the Power 710 and Power 730 to a local area network (LAN), you can use the LAN adapters that are supported in the PCIe slots of the system unit. Table 2-9 lists the additional available LAN adapters.

Table 2-9 Available LAN adapters

Feature code	CCIN	Adapter description	Slot	Size	OS support
5260	576F	PCIe2 LP 4-Port 1 Gb Ethernet adapter	PCIe	Low profile	AIX, IBM i, Linux
5271	5271	PCIe LP 4-Port 10/100/1000 Base-TX Ethernet adapter	PCIe	Low profile	AIX, Linux
5272	5272	PCIe LP 10 Gb Ethernet CX4 1-Port adapter	PCIe	Low profile	AIX, Linux
5274	5768	PCIe LP 2-Port 1 Gb Ethernet SX adapter	PCIe	Low profile	AIX, IBM i, Linux
5275	5275	PCIe LP 10 Gb Ethernet SR 1-Port adapter	PCIe	Low profile	AIX, Linux
5279	2B43	PCIe2 LP 4-Port 10 Gb Ethernet & 1 Gb Ethernet SFP+ Copper & RJ45	PCIe	Low profile	Linux
5280	2B44	PCIe2 LP 4-Port 10 Gb Ethernet & 1 Gb Ethernet SR & RJ45 adapter	PCIe	Low profile	Linux

Feature code	CCIN	Adapter description	Slot	Size	OS support
5281	5767	PCIe LP 2-Port 1 Gb Ethernet TX adapter	PCIe	Low profile	AIX, IBM i, Linux
5284	5287	PCIe2 LP 2-Port 10 Gb Ethernet SR adapter	PCIe	Low profile	AIX, Linux
5286	5288	PCIe2 LP 2-Port 10 Gb Ethernet SFP+ Copper adapter	PCIe	Low profle	AIX, Linux
5717[a]	5271	4-Port 10/100/1000 Base-TX PCI Express adapter	PCIe	Full height	AIX, Linux
5732[a]	5732	10 Gigabit Ethernet-CX4 PCI Express adapter	PCIe	Full height	AIX, Linux
5767[a]	5767	2-Port 10/100/1000 Base-TX Ethernet PCI Express adapter	PCIe	Full height	AIX, IBM i, Linux
5768[a]	5768	2-Port Gigabit Ethernet-SX PCI Express adapter	PCIe	Full height	AIX, IBM i, Linux
5769[a]	5769	10 Gigabit Ethernet-SR PCI Express adapter	PCIe	Full height	AIX, IBM i, Linux
5772[a]	576E	10 Gigabit Ethernet-LR PCI Express adapter	PCIe	Full height	AIX, IBM i, Linux
5899[a]	576F	PCIe2 4-Port 1 Gb Ethernet adapter	PCIe	Full height	AIX, IBM i, Linux
EC27	EC27	PCIe2 LP 2-Port 10 Gb Ethernet RoCE SFP+ adapter	PCIe	Low profile	AIX, Linux
EC29	EC27	PCIe2 LP 2-Port 10 Gb Ethernet RoCE SR adapter	PCIe	Low profile	AIX, Linux
EN0J	2B93	PCIe2 LP 4-Port (10 Gb FCoE & 1 Gb Ethernet) SR & RJ45	PCIe	Low profile	AIX, IBM i, Linux

a. This full height card is supported only in the Power 730 with a FC 5802 or a FC 5877 drawer.

2.7.4 Graphics accelerator adapters

Table 2-10 lists the available graphics accelerator adapters. The adapter can be configured to operate in either 8-bit or 24-bit color modes. The adapter supports both analog and digital monitors.

Table 2-10 Available graphics accelerator adapters

Feature code	CCIN	Adapter description	Slot	Size	OS support
5269	2849	PCIe LP POWER GXT145 Graphics accelerator	PCIe	Low profile	AIX, Linux
5748[a]	5748	POWER GXT145 PCI Express Graphics accelerator	PCIe	Full height	AIX, Linux

a. This full height card is supported only in the Power 730 with a FC 5802 or a FC 5877 drawer.

2.7.5 SAS adapters

Table 2-11 lists the SAS adapter that is available for Power 710 and Power 730 systems.

Table 2-11 Available SAS adapters

Feature code	CCIN	Adapter description	Slot	Size	OS support
5278	58B3	PCIe LP Dual-x4-Port SAS adapter 3 Gb	PCIe	Low profile	AIX, IBM i, Linux
5805[a]	574E	PCIe 380MB Cache Dual - x4 3 Gb SAS RAID adapter	PCIe	Full height	AIX, IBM i, Linux
5901[a]	57B3	PCIe Dual-x4 SAS adapter	PCIe	Full height	AIX, IBM i, Linux
5913[a]	57B5	PCIe2 1.8 GB Cache RAID SAS adapter Tri-port 6 Gb	PCIe	Full height	AIX, IBM i, Linux
ESA1[a]	57B4	PCIe2 RAID SAS adapter Dual-Port 6 Gb	PCIe	Full height	AIX, IBM i, Linux
ESA2	57B4	PCIe2 LP RAID SAS adapter Dual-Port 6 Gb	PCIe	Low profile Short	AIX, IBM i, Linux

a. This full height card is supported only in the Power 730 with an FC 5802 or FC 5877 drawer.

2.7.6 PCIe RAID and SSD SAS adapter

A new SSD option for selected POWER7 and POWER7+ processor-based servers offers a significant price-for-performance improvement for many client SSD configurations. The SSD option is packaged differently from those that are currently available with Power Systems. The PCIe RAID and SSD SAS adapter has up to four 177 GB SSD modules, plugged directly onto the adapter, saving the need for the SAS bays and cabling that are associated with the current SSD offering. This PCIe-based SSD offering can save up to 70% of the list price, and reduce the footprint up to 65%, compared to disk enclosure-based SSD drives, assuming equivalent capacity. This benefit is dependant on the configuration required.

Figure 2-16 shows the double-wide adapter and SSD modules.

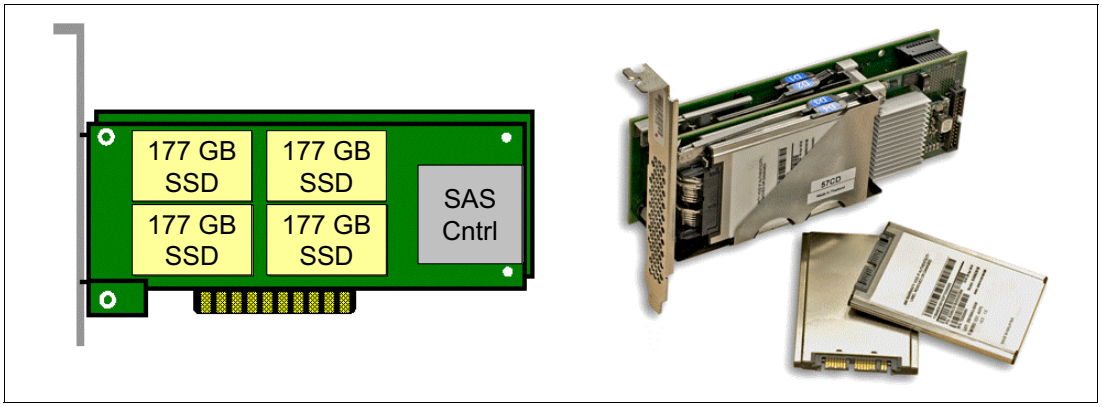


Figure 2-16 The PCIe RAID and SSD SAS adapter, and 177 GB SSD modules

Table 2-12 shows available RAID and SSD SAS adapters for the Power 710 and Power 730.

Table 2-12 Available PCIe RAID and SSD SAS adapters

Feature code	CCIN	Adapter description	Slot	Size	OS support
2053	57CD	PCIe LP RAID & SSD SAS adapter 3 Gb	PCIe	Low profile double-wide, short	AIX, IBM i, Linux
2055[a]	57CD	PCIe RAID & SSD SAS adapter 3 Gb with Blind Swap Cassette	PCIe	Full height	AIX, IBM i, Linux

a. This full height card is supported only in the Power 730 with an FC 5802 or FC 5877 drawer.

The 177 GB SSD module with enterprise multi-level cell (eMLC) uses an enterprise-class MLC flash technology, which provides enhanced durability, capacity, and performance. One, two, or four modules can be plugged onto a PCIe RAID and SSD SAS adapter, providing up to 708 GB of SSD capacity on one PCIe adapter.

Because the SSD modules are mounted on the adapter, to service either the adapter or one of the modules, the entire adapter must be removed from the system.

Under AIX and Linux, the 177 GB modules can be reformatted as JBOD disks, providing 200 GB of available disk space. This way removes RAID error correcting information, so the best approach to prevent data loss in case of failure is to mirror the data by using operating system tools.

2.7.7 Fibre Channel adapters

The systems support direct or SAN connection to devices that use Fibre Channel adapters. Table 2-13 summarizes the available Fibre Channel adapters, which all 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 (FC 2456) or an LC-SC 62.5 Micron Fiber Converter Cable (FC 2459) is required.

Table 2-13 Available Fibre Channel adapters

Feature code	CCIN	Adapter description	Slot	Size	OS support
5273	577D	PCIe LP 8 Gb 2-Port Fibre Channel adapter	PCIe	Low profile Short	AIX, IBM i, Linux
5276	5774	PCIe LP 4 Gb 2-Port Fibre Channel adapter	PCIe	Low profile Short	AIX, IBM i, Linux
5735[a]	577D	8 Gigabit PCI Express Dual Port Fibre Channel adapter	PCIe	Full height Short	AIX, IBM i, Linux
5773[a]	5773	4 Gigabit PCI Express Single Port Fibre Channel adapter	PCIe	Full height	AIX, Linux
5774[a]	5774	4 Gigabit PCI Express Dual Port Fibre Channel adapter	PCIe	Full height	AIX, Linux

Feature code	CCIN	Adapter description	Slot	Size	OS support
EN0Y	EN0Y	PCIe2 LP 8 Gb 4-Port Fibre Channel adapter	PCIe	Low profile Short	AIX, IBM i, Linux
EN0B	577D	PCIe2 LP 16 Gb 2-Port Fibre Channel adapter	PCIe	Low profile	AIX, IBM i, Linux

a. This full height card is only supported in the Power 730 with a FC 5802 or a FC 5877 drawer.

NPIV: The use of N_Port ID Virtualization (NPIV) through the Virtual I/O server requires an NPIV-capable Fibre Channel adapter, such as the FC 5273, FC 5735, and FC EN0B.

2.7.8 Fibre Channel over Ethernet

Fibre Channel over Ethernet (FCoE) allows for the convergence of Fibre Channel and Ethernet traffic onto a single adapter and a converged fabric.

Figure 2-17 compares existing Fibre Channel and network connections and FCoE connections.

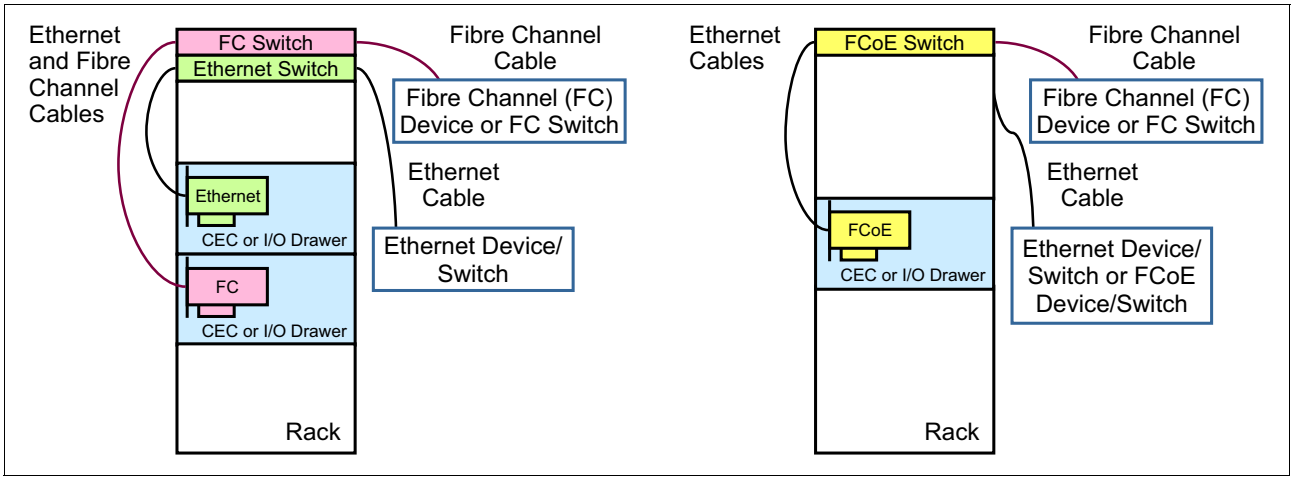


Figure 2-17 Comparison between existing Fibre Channel and network connection and FCoE connection

Table 2-14 lists the available Fibre Channel over Ethernet Adapter. It is a high-performance, Converged Network Adapter (CNA) using SR optics. Each port can simultaneously provide network interface card (NIC) traffic and Fibre Channel functions.

Table 2-14 Available FCoE adapters

Feature code	CCIN	Adapter description	Slot	Size	OS support
5270	2B3B	PCIe LP 10 Gb FCoE 2-Port adapter	PCIe	Low profile Short	AIX, Linux
EN0J	2B93	PCIe2 LP 4-Port (10 Gb FCoE & 1 Gb Ethernet) SR & RJ45 adapter	PCIe	Low profile	AIX, IBM i, Linux

For more information about FCoE, read *An Introduction to Fibre Channel over Ethernet, and Fibre Channel over Convergence Enhanced Ethernet*, REDP-4493.

2.7.9 InfiniBand Host Channel adapter

The InfiniBand Architecture (IBA) is an industry-standard architecture for server I/O and inter-server communication. It was developed by the InfiniBand Trade Association (IBTA) to provide the levels of reliability, availability, performance, and scalability necessary for present and future server systems with levels significantly better than can be achieved using bus-oriented I/O structures.

InfiniBand (IB) is an open set of interconnect standards and specifications. The main IB specification is published by the InfiniBand Trade Association and is available at:

http://www.infinibandta.org/

InfiniBand is based on a switched fabric architecture of serial point-to-point links, where these IB links can be connected to either host channel adapters (HCAs), used primarily in servers, or target channel adapters (TCAs), used primarily in storage subsystems.

The InfiniBand physical connection consists of multiple byte lanes. Each individual byte lane is a four-wire, 2.5, 5.0, or 10.0 Gbps bidirectional connection. Combinations of link width and byte lane speed allow for overall link speeds from 2.5 Gbps to 120 Gbps. The architecture defines a layered hardware protocol and also a software layer to manage initialization and the communication between devices. Each link can support multiple transport services for reliability and multiple prioritized virtual communication channels.

For more information about InfiniBand, see *HPC Clusters Using InfiniBand on IBM Power Systems Servers*, SG24-7767.

The GX++ Dual-port 12X Channel Attach adapter (FC EJ0G) provides two 12X connections for 12X Channel applications. One adapter must be installed in GX++ bus slot 2 and will cover one adjacent PCIe x8 G2 slot 5. The 12X Channel is connected in a loop and uses both connectors on the adapters. Up to two I/O drawers can be attached in a single loop. This adapter must be used with the 12X cables.

A connection to supported InfiniBand switches is accomplished by using the 12x to 4x Channel Conversion Cables, FC 1828, FC 1841, or FC 1842.

Table 2-15 lists the available InfiniBand adapters.

Table 2-15 Available InfiniBand adapters

Feature code	CCIN	Adapter description	Slot	Size	OS support
5283		PCIe2 LP Dual-Port 4X IB QDR adapter 40 Gb	PCIe	Low profile Short	AIX, Linux
EJ0G[a]		GX++ Dual-Port 12X Channel Attach adapter	GX++ PCIe		AIX, Linux

a. FC EJ0G is not available for the Power 710 server.

2.7.10 Asynchronous and USB adapters

Asynchronous PCIe 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 Power Systems using a serial connection, you can use the features listed in Table 2-16.

Table 2-16 Available asynchronous adapters

Feature code	CCIN	Adapter description	Slot	Size	OS support
2728[a]	57D1	4-Port USB PCIe adapter	PCIe	Full height	AIX, Linux
5277	57D2	PCIe LP 4-Port Async EIA-232 adapter	PCIe	Low profile Short	AIX, Linux
5290	57D4	PCIe LP 2-Port Async EIA-232 adapter	PCIe	Low profile Short	AIX, Linux
5785[a]	57D2	4-Port Async EIA-232 PCIe adapter	PCIe	Full height	AIX

a. This full height card is supported only in the Power 730 with a FC 5802 or a FC 5877 drawer.

2.7.11 Cryptographic coprocessor

The cryptographic coprocessor cards provide both cryptographic coprocessor and cryptographic accelerator functions in a single card.

The IBM PCIe Cryptographic Coprocessor adapter highlights the following features:

- Integrated Dual processors that operate in parallel for higher reliability
- Supports IBM Common Cryptographic Architecture or PKCS#11 standard
- Ability to configure adapter as coprocessor or accelerator
- Support for smart card applications that use Europay, MasterCard and Visa
- Cryptographic key generation and random number generation
- PIN processing; generation, verification, translation
- Encrypt and decrypt by using AES and DES keys

See the following location for the latest firmware and software updates:

`http://www.ibm.com/security/cryptocards/`

Table 2-17 lists the cryptographic adapter that is available for the server.

Table 2-17 Available cryptographic adapters

Feature code	CCIN	Adapter description	Slot	Size	OS support
4808[a]	4765	PCIe Crypto Coprocessor with GEN3 Blindswap Cassette 4765-001	PCIe	Full height	AIX, IBM i, Linux

a. This full height card is supported only in the Power 730 with a FC 5802 or a FC 5877 drawer.

2.8 Internal storage

The Power 710 and Power 730 servers use an integrated SAS/SATA controller connected through a PCIe bus to the P7IOC chip supporting RAID 0, 1, and 10 (see Figure 2-18). The SAS/SATA controller in the server's enclosure has two sets of four SAS/SATA channels, which give the Power 710 and Power 730 systems the combined total of eight SAS buses. Each channel can support either SAS or SATA operation. The SAS controller is connected to a DASD backplane and supports three or six small form factor (SFF) disk drive bays, depending on the backplane option.

One of the following options must be selected as the backplane:

- FC EJ0E supports three SFF disk units, either HDD or SSD, an SATA DVD, and a tape (FC 5762 or follow on). There is no support for split backplane and RAID 5 and 6.

  IBM i: Feature FC EJ0E is not supported with IBM i.

- FC EJ0D supports six SFF disk units, either HDD or SSD, and a SATA DVD. There is no support for split backplane and RAID 5 and 6.
- FC EJ0F supports six SFF disk units, either HDD or SSD, a SATA DVD, a Dual Write Cache RAID, and an external SAS port. HDDs/SSDs are hot-swap and front accessible. Split backplane is not supported. RAID levels 5 and 6 are supported. This feature is required when IBM i is the primary operating system (FC 2145).

The supported disk drives in a Power 710 and Power 730 server connect to the DASD backplane and are hot-swap and front-accessible.

Figure 2-18 details the internal topology overview for the FC EJ0E backplane.

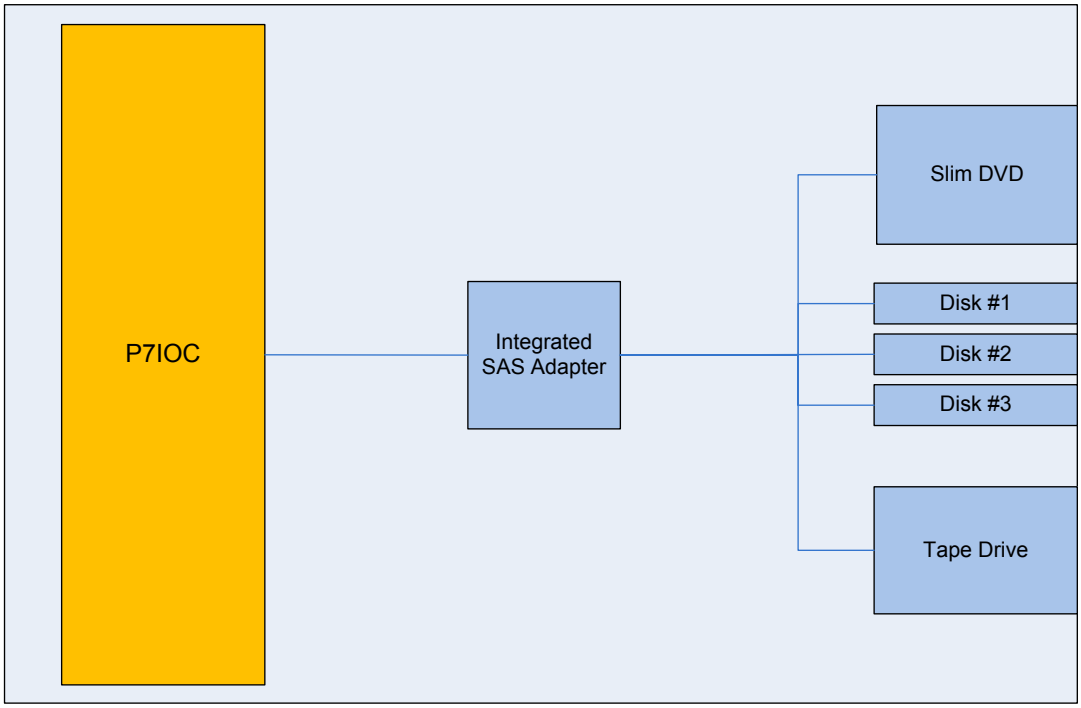


Figure 2-18 Internal topology overview for FC EJ0E DASD backplane

Figure 2-19 shows the internal topology overview for the FC EJ0D backplane.

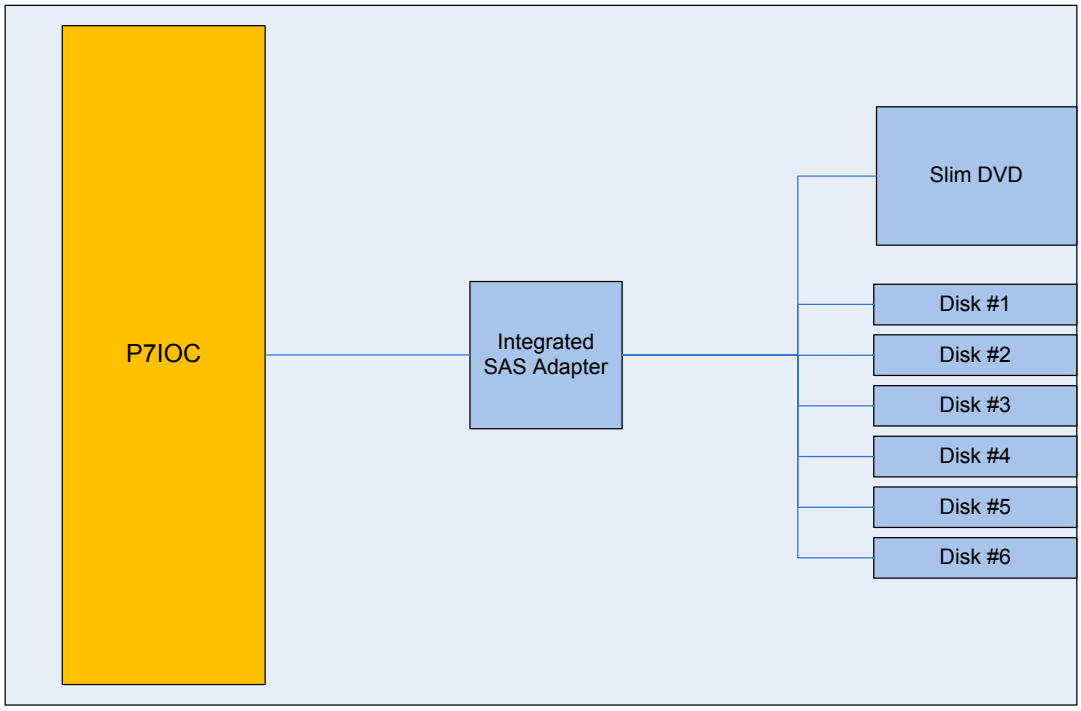


Figure 2-19 Internal topology overview for the FC EJ0D DASD backplane

Figure 2-20 shows the details of the internal topology overview for the FC 5268 backplane.

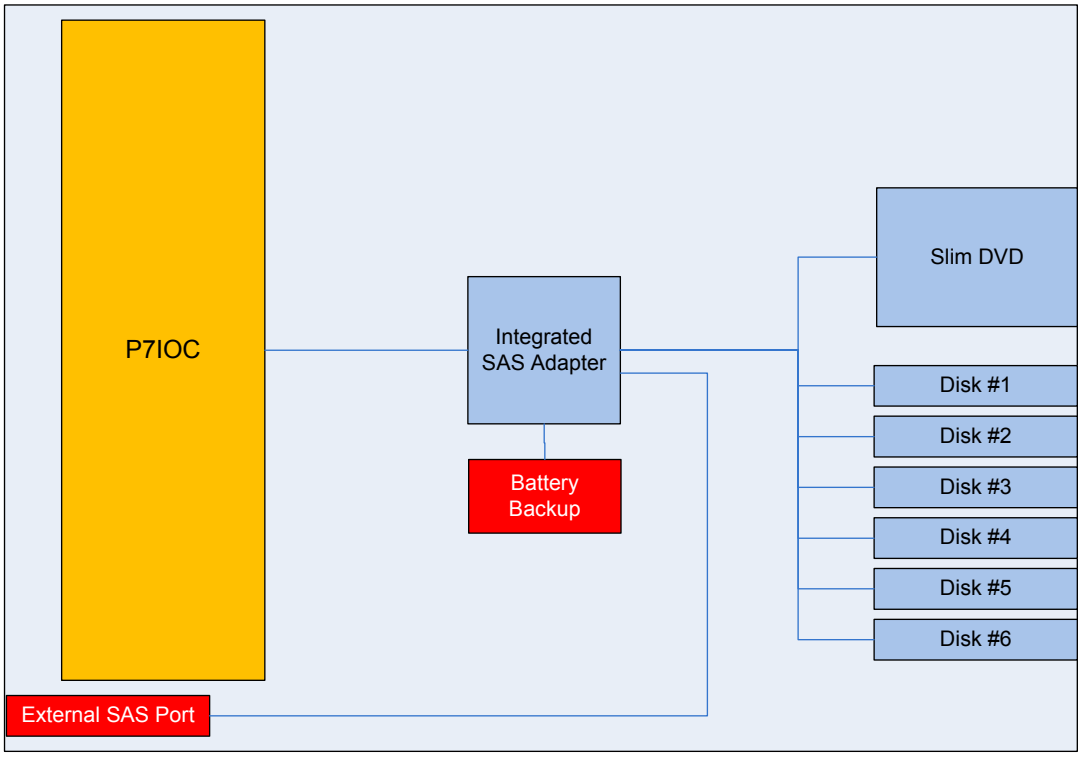


Figure 2-20 Internal topology overview for the FC EJ0F DASD backplane

2.8.1 RAID support

There are multiple protection options for HDD/SSD drives in the Power 710 and Power 730 systems, whether they are contained in the SAS SFF bays in the system unit, in a 12X attached I/O drawer, or drives in disk-only I/O drawers. Although protecting drives is always recommended, AIX/Linux users can choose to leave a few or all drives unprotected at their own risk, and IBM supports these configurations. IBM i configuration rules differ in this regard, and IBM supports IBM i partition configurations only when HDD/SSD drives are protected.

Drive protection

HDD/SSD drive protection can be provided by the AIX, IBM i, and Linux operating system, or by the HDD/SSD hardware controllers. Mirroring of drives is provided by the AIX, IBM i, and Linux operating system. In addition, AIX and Linux support controllers providing RAID 0, 1, 5, 6, or 10. The integrated storage management of IBM i already provides striping. IBM i also supports controllers that provide RAID 5 or 6. To further augment HDD/SSD protection, hot-spare capability can be used for protected drives. Specific hot-spare prerequisites apply.

An integrated SAS HDD/SSD controller is in the Power 710 and Power 730 system unit and provides support for JBOD and RAID 0, 1, and 10 for AIX or Linux. It is optionally augmented by RAID 5 and RAID 6 capability when storage backplane FC EJ0F is added to the configuration. In addition to these protection options, mirroring of drives by the operating system is supported. AIX or Linux supports all of these options. IBM i does not use unprotected disks and uses embedded functions instead of RAID 10. IBM i does use the RAID 5 or 6 function of the integrated controllers.

Table 2-18 lists the RAID support by backplane.

Table 2-18 RAID support configurations

Storage backplane	JBOD	RAID 0, 1, and 10	RAID 5 and 6	Split backplane	External SAS port
FC EJ0D	Yes	Yes	No	No	No
FC EJ0E	Yes	Yes	No	No	No
FC EJ0F	No	Yes	Yes	No	Yes

AIX and Linux can use disk drives that are formatted with 512-byte blocks when being mirrored by the operating system. These disk drives must be reformatted to 528-byte sectors when used in RAID arrays. Although a small percentage of the drive's capacity is lost, additional data protection such as ECC and bad block detection is gained in this reformatting. For example, a 300 GB disk drive, when reformatted, provides approximately 283 GB. IBM i always uses drives that are formatted to 528 bytes. Solid-state drives are always formatted with 528 byte sectors.

The Power 710 and Power 730 support a dual-write cache RAID feature that consists of an auxiliary write cache for the RAID card and the optional RAID enablement.

Supported RAID functions

Base hardware supports RAID 0, 1, and 10. When additional features are configured, Power 710 and Power 730 support hardware RAID 0, 1, 5, 6, and 10:

- RAID 0 provides striping for performance, but does not offer any fault tolerance.

  The failure of a single drive results in the loss of all data on the array. This version of RAID increases I/O bandwidth by simultaneously accessing multiple data paths.

- RAID 1 mirrors the contents of the disks. The contents of each disk in the array are identical to that of every other disk in the array, providing data resilience in the case of a drive failure.

- RAID 5 uses block-level data striping with distributed parity.

  RAID RAID 5 stripes both data and parity information across three or more drives. Fault tolerance is maintained by ensuring that the parity information for any given block of data is placed on a drive separate from those used to store the data itself. This version of RAID provides data resiliency in the case of a single drive failing in a RAID 5 array.

- RAID 6 uses block-level data striping with dual distributed parity.

  RAID 6 is the same as RAID 5 except that it uses a second level of independently calculated and distributed parity information for additional fault tolerance. RAID 6 configuration requires N+2 drives to accommodate the additional parity data, making it less cost effective than RAID 5 for equivalent storage capacity. This version of RAID provides data resiliency in the case of one or two drives failing in a RAID 6 array.

- RAID 10 is also known as a striped set of mirrored arrays.

  It is a combination of RAID 0 and RAID 1. A RAID 0 stripe set of the data is created across a two-disk array for performance benefits. A duplicate of the first stripe set is then mirrored on another two-disk array for fault tolerance. This version of RAID provides data resiliency in the case of a single drive failure and may provide resiliency for multiple drive failures.

2.8.2 External SAS port

The Power 710 and Power 730 DASD backplane (FC EJ0F) offers a connection to an external SAS port:

- The SAS port connector is located next to the GX++ slot 2 on the rear bulkhead.
- The external SAS port is used for expansion to external SAS devices or drawer such as the EXP12S SAS Drawer (FC 5886), the EXP24S SFF Gen2-bay Drawer (FC 5887), and the IBM System Storage 7214 Tape and DVD Enclosure Express (Model 1U2).

Note: Only one SAS drawer is supported from the external SAS port. Additional SAS drawers can be supported through SAS adapters. SSDs are not supported on the SAS drawer connected to the external port.

2.8.3 Media bays

The Power 710 and Power 730 each have a slim media bay that contains an optional DVD-RAM (FC 5762) and a tape bay (available only with FC EJ0E) that can contain a tape drive or removable disk drive. Direct dock and hot-plug of the DVD media device is supported.

The DVD drive and media device do not have an independent SAS adapter and so cannot be assigned to an LPAR independently of the HDD/SSDs in the system.

2.9 External I/O subsystems

The Power 730 server supports the attachment of I/O drawers. Two of the following I/O drawers can be attached to the system unit, providing extensive capability to expand the overall server:

- 12X I/O Drawer PCIe, small form factor (SFF) disk (FC 5802)
- 12X I/O Drawer PCIe, No Disk (FC 5877)

Each processor card feeds one GX++ adapter slot. Two GX++ slots are available in the Power 730.

The Power 730 uses the GX++ Dual-port 12x Channel Attach (FC EJ0G) adapter to attach a FC 5802 or FC 5877 12X I/O Drawer. The FC EJ0G provides double data rate (DDR) capacity bandwidth.

Table 2-19 is an overview of the capabilities of the supported I/O drawers.

Table 2-19 I/O drawer capabilities

Feature code	Disk drive bays	PCI slots	Requirements for Power 710 and Power 730
5802	18 SAS hot-swap disk drive bays	10 PCIe	GX++ Dual-port 12x Channel Attach (FC EJ0G)
5877	None	10 PCIe	GX++ Dual-port 12x Channel Attach (FC EJ0G)

Unsupported: The attachment of external I/O drawers is not supported on the Power 710.

2.9.1 12X I/O Drawer PCIe

The 12X I/O Drawer PCIe, SFF disk (FC 5802) is a 19-inch I/O and storage drawer. It provides a 4U (EIA units) drawer, containing 10 PCIe-based I/O adapter slots and 18 SAS hot-swap small form factor (SFF) disk bays, which can be used for either disk drives or SSD drives. Using 900 GB disk drives, each I/O drawer provides up to 16.2 TB of storage. The adapter slots within the I/O drawer use Gen3 blind swap cassettes and support hot-plugging of adapter cards. The 12X I/O Drawer PCIe, No Disk (FC 5877) is the same as FC 5802 except that it does not support any disk bays.

A maximum of two 12X I/O Drawer PCIe, SFF disk drawers can be placed on the same 12X loop. Within the same loop, FC 5877 and FC 5802 can be mixed. An upgrade from a diskless FC 5877 to FC 5802 with disk bays is not available.

A minimum configuration of two 12X DDR cables, two AC power cables, and two SPCN cables is required to ensure proper redundancy. The drawer attaches to the system unit with a 12X adapter in a GX++ slot through 12X DDR cables that are available in the following cable lengths:

- 0.6 meters 12X DDR Cable (FC 1861)
- 1.5 meters 12X DDR Cable (FC 1862)
- 3.0 meters 12X DDR Cable (FC 1865)
- 8.0 meters 12X DDR Cable (FC 1864)

Unsupported: The 12X SDR cables are not supported.

The physical dimensions of the drawer measure 444.5 mm (17.5 in.) wide, by 177.8 mm (7.0 in.) high, by 711.2 mm (28.0 in.) deep for use in a 19-inch rack.

Figure 2-21 shows the front view of the 12X I/O Drawer PCIe (FC 5802).

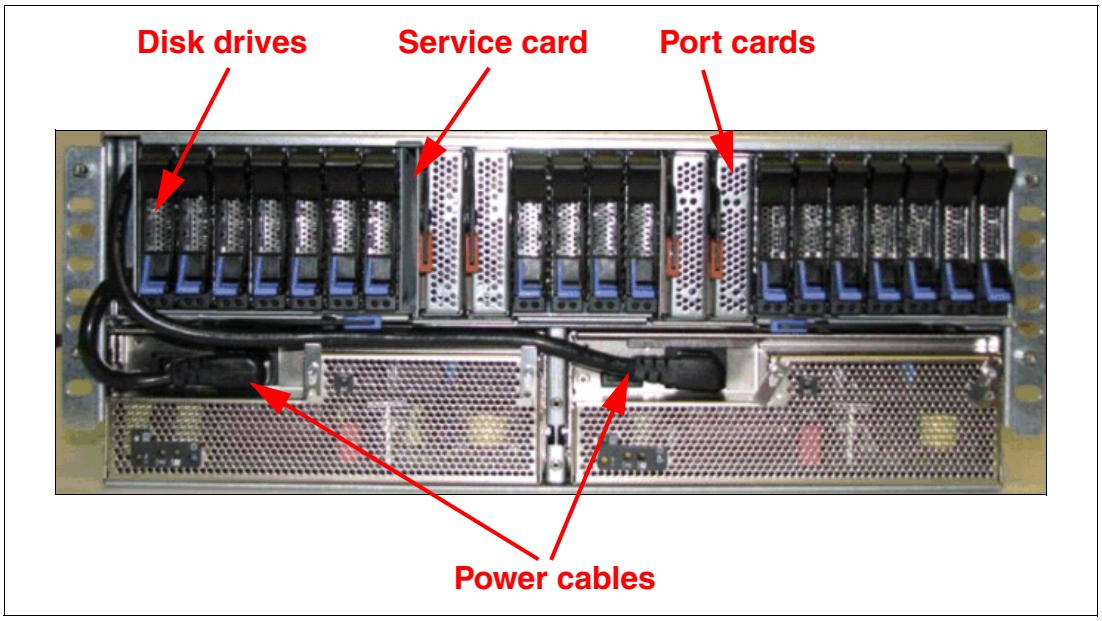


Figure 2-21 Front view of the 12X I/O Drawer PCIe

Figure 2-22 shows the rear view of the 12X I/O Drawer PCIe (FC 5802).

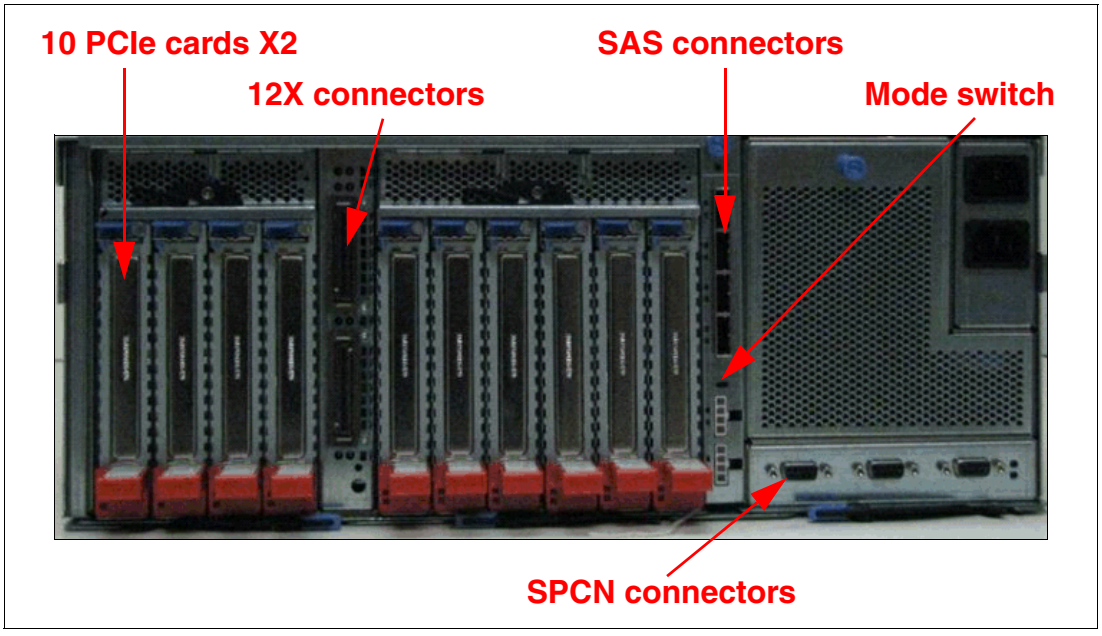


Figure 2-22 Rear view of the 12X I/O Drawer PCIe

2.9.2 12X I/O Drawer PCIe configuration and cabling rules

The following sections describe the disk drive configuration, 12X loop, and SPCN cabling rules.

Configuring the disk drive subsystem of the FC 5802 drawer

The 12X I/O Drawer PCIe, SFF disk drawer (FC 5802) can hold up 18 disk drives. The disks in this enclosure can be organized in various configurations depending on the operating system used, the type of SAS adapter card, and the position of the mode switch.

Each disk bay set can be attached to its own controller or adapter. Feature PCIe 12X I/O drawer has four SAS connections to drive bays. It connects to PCIe SAS adapters or controllers on the host systems.

Disk drive bays in the 12X I/O drawer PCIe can be configured as one, two, or four sets. This way allows for partitioning of disk bays. Disk bay partitioning configuration can be done with the physical mode switch on the I/O drawer.

Remember: A mode change, using the physical mode switch, requires the drawer to be powered off and then on.

Figure 2-23 indicates the mode switch in the rear view of the FC 5802 I/O Drawer and shows the configuration rules of disk bay partitioning in the PCIe 12X I/O drawer. There is no specific feature code for mode switch setting.

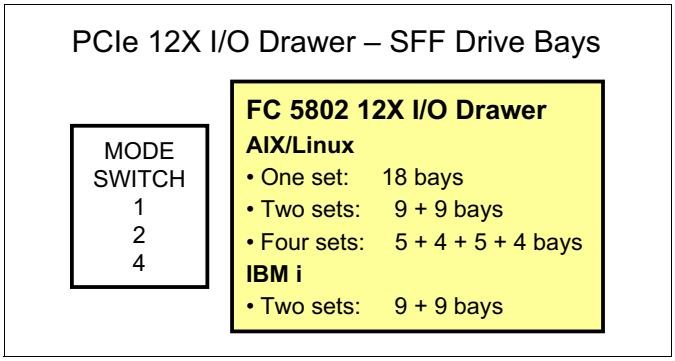


Figure 2-23 Disk bay partitioning configuration in 12X I/O Drawer PCI (FC 5802)

Tools and CSP: The IBM System Planning Tool supports disk bay partitioning. Also, the IBM configuration tool accepts this configuration from IBM System Planning Tool and passes it through IBM manufacturing by using the Customer Specified Placement (CSP) option.

The location codes for the front and rear views of the FC 5802 I/O drawer are provided in Figure 2-24 and Figure 2-25.

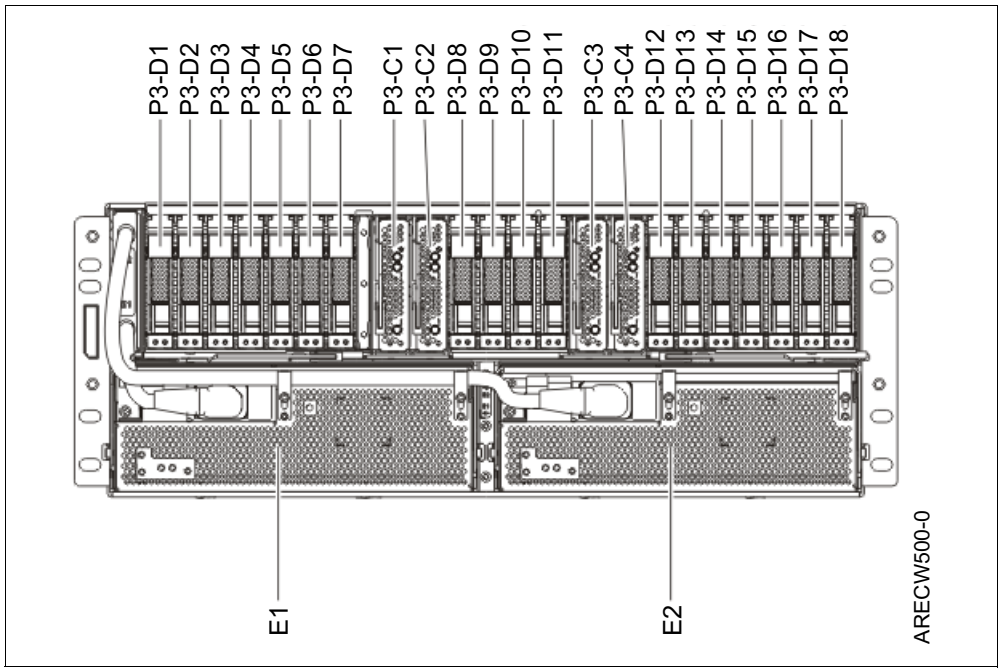


Figure 2-24 FC 5802 I/O drawer from view location codes

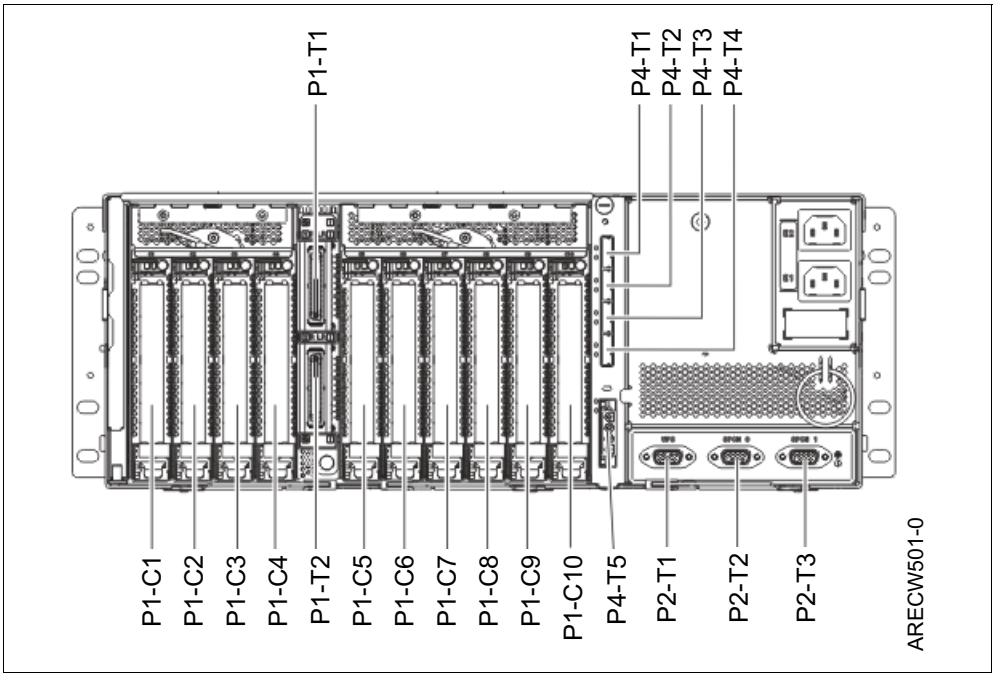


Figure 2-25 FC 5802 I/O drawer rear view location codes

Table 2-20 lists the SAS ports that are associated to the disk bays with the mode selector switch 4.

Table 2-20 SAS connection mappings

Location code	Mappings	Number of bays
P4-T1	P3-D1 to P3-D5	5 bays
P4-T2	P3-D6 to P3-D9	4 bays
P4-T3	P3-D10 to P3-D14	5 bays
P4-T4	P3-D15 to P3-D18	4 bays

For more detailed information about cabling and other switch modes, see the *Power Systems Enclosures and expansion units* publication:

http://pic.dhe.ibm.com/infocenter/powersys/v3r1m5/topic/ipham/ipham.pdf

12X I/O Drawer PCIe loop

Any I/O drawer is connected to the adapters in the Power 730 system unit with data transfer cables such as the 12X DDR cables for the FC 5802 and FC 5877 I/O drawers.

The first 12X I/O drawer that is attached to the I/O drawer loop requires two data transfer cables. An additional second drawer requires one additional data transfer cable. Consider the following information about the loop:

- A 12X I/O loop starts at a system unit adapter port 0 and attaches to port 0 of the first I/O drawer.
- The I/O drawer attaches from port 1 of the first unit to port 0 of the second I/O drawer.
- Port 1 of the second I/O drawer on the 12X I/O loop connects to port 1 of the system unit adapter to complete the loop.

Figure 2-26 shows typical 12X I/O loop port connections.

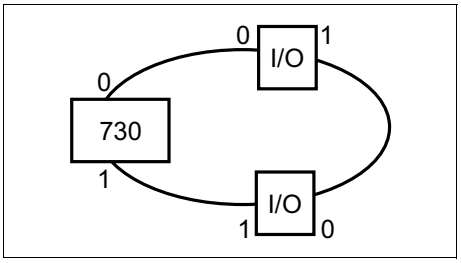


Figure 2-26 Typical 12X I/O loop port connections

Table 2-21 lists 12X cables to satisfy the various length requirements.

Table 2-21 12X connection cables

Feature code	Description
1861	0.6 meter 12X DDR cable
1862	1.5 meter 12X DDR cable
1865	3.0 meter 12X DDR cable
1864	8.0 meter 12X DDR cable

12X I/O Drawer PCIe SPCN cabling

System Power Control Network (SPCN) is used to control and monitor the status of power and cooling within the I/O drawer.

SPCN cables connect all AC powered expansion units. Figure 2-27 shows an example for a Power 730 that connects to two I/O drawers. Other connections options are available.

1. Start at SPCN 0 (T1) of the Power 730 system unit to J15 (T1) of the first I/O drawer.
2. Cable all units from J16 (T2) of the first I/O drawer to J15 (T1) of the second I/O drawer.
3. To complete the cabling loop, from J16 (T2) of the second I/O drawer, connect to the system unit SPCN 1 (T2).
4. Ensure that a complete loop exists from the system unit, through all attached expansions and back to the system unit.

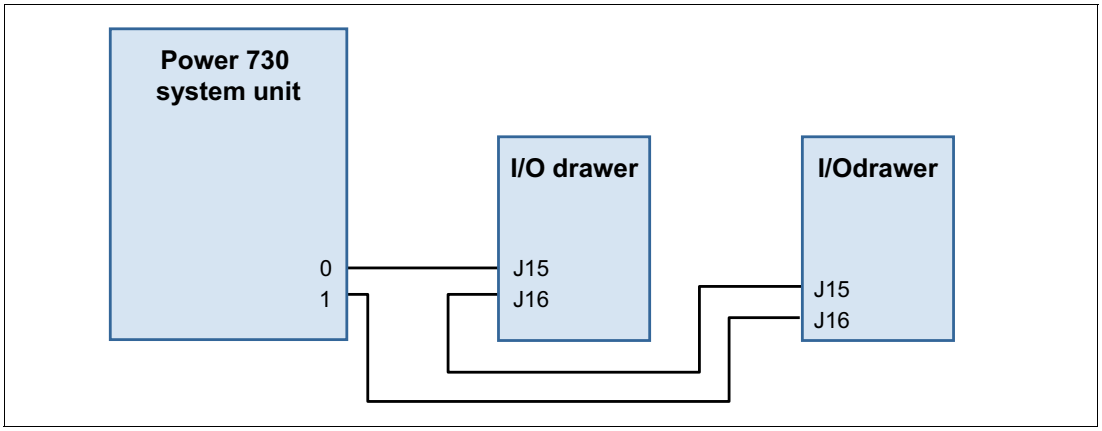


Figure 2-27 SPCN cabling example

Various SPCN cables are available. Table 2-22 lists available SPCN cables options to satisfy various length requirements.

Table 2-22 SPCN cables

Feature code	Description
6001[a]	Power Control Cable (SPCN) - 2 meter
6006	Power Control Cable (SPCN) - 3 meter
6008[a]	Power Control Cable (SPCN) - 6 meter
6007	Power Control Cable (SPCN) - 15 meter
6029[a]	Power Control Cable (SPCN) - 30 meter

a. Supported, but no longer orderable

2.10 External disk subsystems

This section describes the following external disk subsystems that can be attached to the Power 710 and Power 730:

- EXP30 Ultra SSD I/O drawer (FC EDR1, CCIN 57C3)
- EXP24S SFF Gen2-bay drawer for high-density storage (FC 5887)
- EXP12S SAS expansion drawer (FC 5886)
- IBM System Storage

Later sections give you detailed information about the various external disk subsystems.

2.10.1 EXP30 Ultra SSD I/O drawer

The EXP30 Ultra SSD I/O drawer (FC EDR1) is a 1U high I/O drawer that provides 30 hot-swap SSD bays and a pair of integrated large write caches, high-performance SAS controllers without using any PCIe slots on the POWER7+ server. The two high performance, integrated SAS controllers each physically provide 3.1 GB write cache. Working as a pair, they provide mirrored write-cache data and controller redundancy. The cache contents are designed to be protected by built-in flash memory in case of power failure. If the pairing is broken, write cache is not used after existing cache content is written out to the drive, and performance will probably be slowed until the controller pairing is established again.

Figure 2-28 shows the front view of the EXP30 drawer.

Figure 2-28 Front view of the EXP30 Ultra SSD I/O drawer

Each controller is connected to a GX++ LP 1-port PCIe2 x8 adapter (FC EJ0H CCIN 2C1F) in a Power 710 and Power 730 server over a PCIe x8 cable. Usually both controllers are attached to one server, but each controller can be assigned to a separate server, or a logical partition.

Table 2-23 lists the RAID levels for the AIX, IBM i, Linux operating systems that the controller supports.

Table 2-23 Supported RAID levels

RAID level	Operating system
RAID 0	AIX, Linux
RAID 1[a]	AIX, IBM i, Linux
RAID 5	AIX, IBM i, Linux
RAID 6	AIX, IBM i, Linux
RAID 10	AIX, Linux

a. Provided by the operating system Logical Volume Manager (LVM)

The EXP30 Ultra SSD I/O drawer (FC EDR1) delivers up to 480,000 IOPS (read only), up to 410,000 IOPS (60% read and 40% write), or up to 325,000 IOPS (100% write) and has up to 30% performance improvement over the previous version of the EXP30 (FC 5888).

Table 2-24 lists the quantity of EXP30 drawers that can be attached to the Power 710 and Power 730, running separate operating systems.

Table 2-24 Quantity of EXP30 attachments

System	AIX	IBM i	Linux
Power 710	One half of the EXP30 drawer	-	One half of the EXP30 drawer
Power 730	One	One[a]	One

a. At the time of writing, only one EXP30 drawer is supported when using the IBM i operating system

Disks

The 387 GB SSD (FC ES02 and FC ES04) used in the EXP30 Ultra SSD I/O drawer uses high-performance, industrial-strength eMLC technology. These SSDs are packaged as 1.8-inch SAS drives, which can be added to or removed concurrently while the drawer is in use.

A minimum of six SSDs are required in each Ultra drawer. Each controller can access all 30 SSD bays. The bays can be configured as one set of bays that is run by a pair of controllers that are working together. Alternatively the bays can be divided into two logical sets, where each of the two controllers owns one of the logical sets. With proper software, if one of the controller fails, the other controller can run both sets of bays.

FC ES02 and FC ES04 are identical SSD drives, but have separate feature codes for use with the AIX, IBM i, and Linux operating systems. FC ES02 is used for AIX and Linux; FC ES04 is used for IBM i.

Unsupported: The 387 GB 1.8" SAS SSD for IBM i with eMLC (FC ES04) is not supported with the Power 710.

EXP30 connection to a Power Systems server

The GX++ LP 1-Port PCIe2 x8 adapter (FC EJ0H, CCIN 2C1F) enables the attachment of the EXP30 Ultra SSD I/O Drawer. The adapter is plugged into a GX++ slot of the 2U Power 710 or Power 730. Up to one PCIe cable connects the drawer to the GX++ 1-Port adapter.

The following cable lengths are available to connect a drawer with a GX++ LP 1-Port PCIe2 x8 adapter.

- 1.5 meters (FC EN05)
- 3 meters (FC EN07)

When connecting one EXP30 drawer to a Power 710 server, only one half of the EXP30 can be connected to the GX++ LP 1-Port PCIe2 x8 adapter. The second half of the EXP30 must be connected to another GX++ LP 1-Port PCIe2 x8 adapter in a different server: either a Power 710, Power 720, Power 730, Power 740, Power 750, Power 760, Power 770, or Power 780 server.

Figure 2-29 shows two examples for a supported solution. The top example shows connecting both Power 710 systems to an EXP30 drawer; the bottom example shows connecting a Power 710 and a Power 750 to an EXP30 drawer.

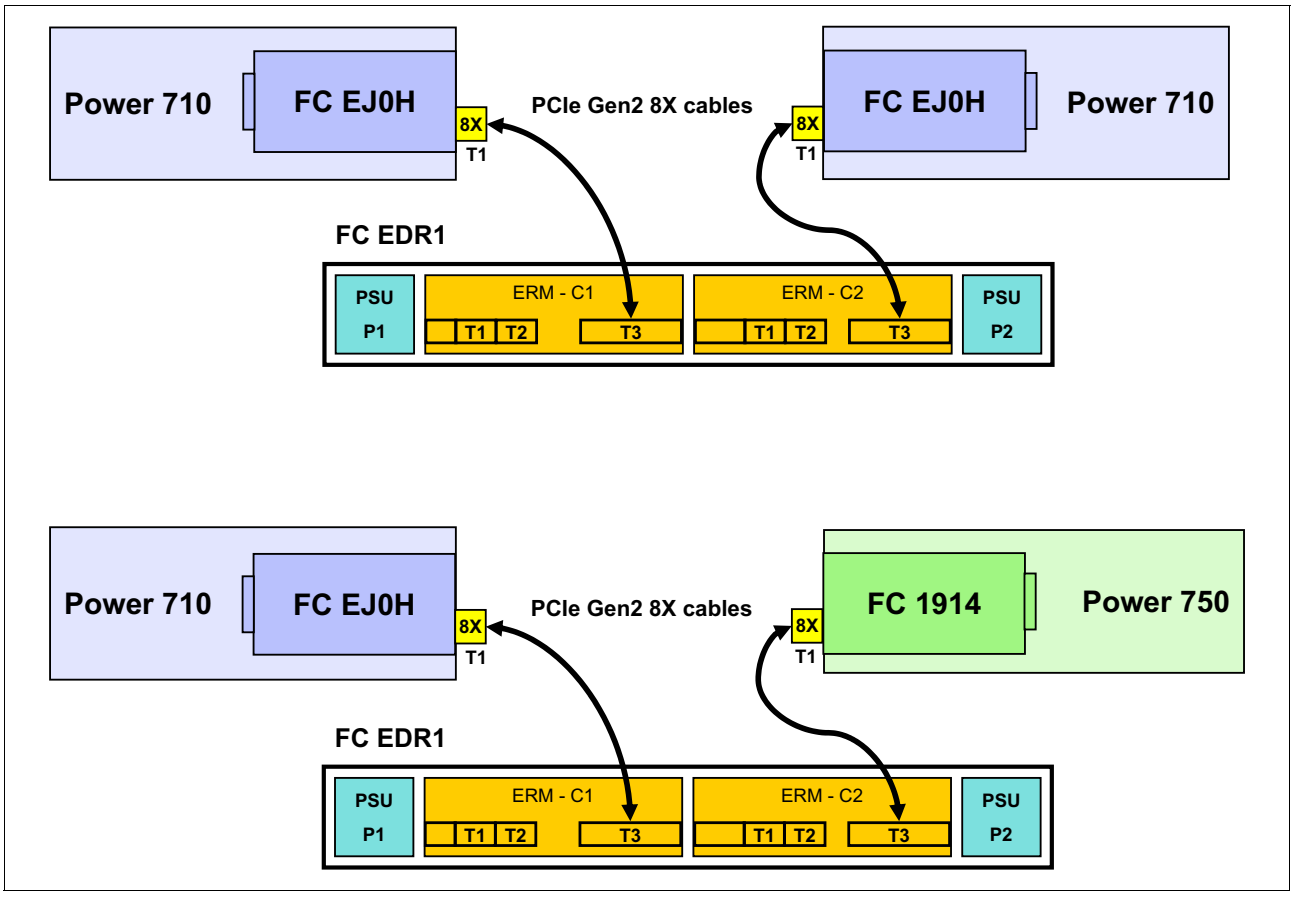


Figure 2-29 Connection between Power 710, an FC EDR1 drawer and a Power 750 server

Figure 2-30 shows a Power 730, using two GX++ LP 1-Port PCIe2 x8 adapters, connected to an EXP30 drawer.

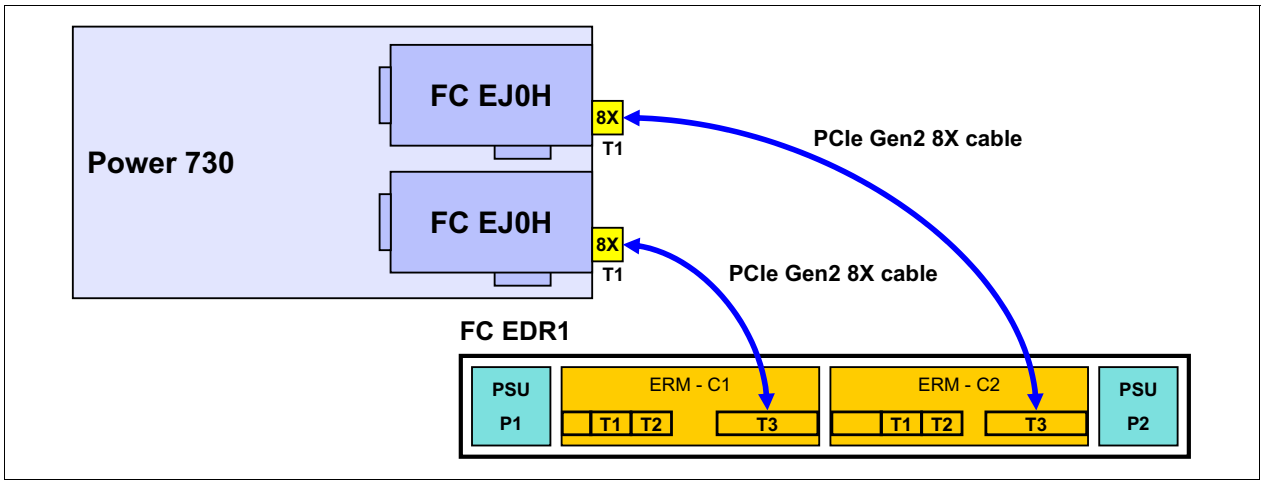


Figure 2-30 Connection to EXP30 drawer

EXP30 drawer connection to EXP24S drawer

Two EXP24S disk drawers (FC 5887) can be directly attached to an EXP30 (FC EDR1) drawer, running AIX, IBM i, and Linux. Up to 48 additional SAS disks enhance the disk capacity up to 43.2 TB. This combination (one EXP30 Ultra Drawer and two EXP24S drawers) provides a maximum capacity of 54.8 TB capacity.

Use both T1 connector locations of the EXP30 drawer to connect an EX SAS cable to the two T1 connector locations of the first EXP24S drawer. If you want to attach a second EXP24S drawer, connect both T2 connector locations of the EXP30 drawer with the two T1 connector locations of the second EXP24S drawer.

Support:

- IBM i 7.1 TR6 also supports attaching downstream EXP24S drives, but has a maximum of one downstream EXP24S drawer and therefore a maximum of up to 24 additional SAS disks.
- The previous model of the EXP30 drawer (FC 5888) does not support the attachment of an EXP24S drawer.

Figure 2-31 shows two EXP24S drawers that are connected to one EXP30 drawer.

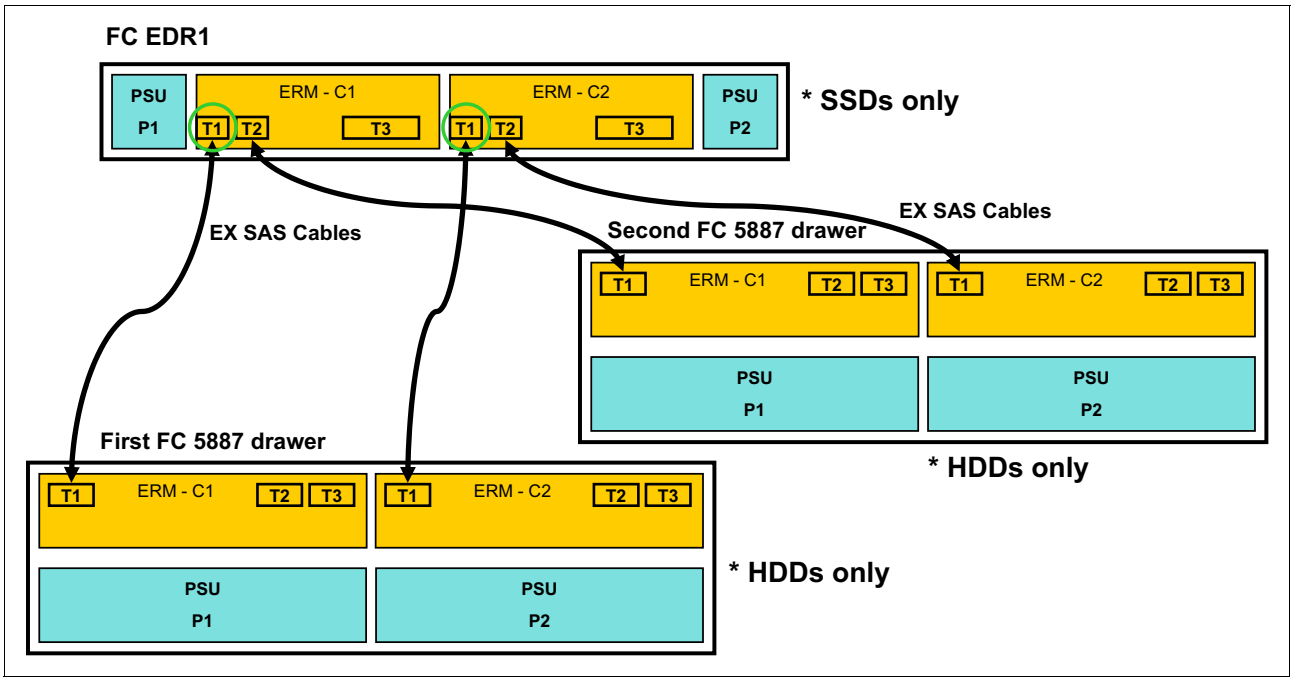


Figure 2-31 FC EDR1 drawer connection to two FC 5887 drawers

More information about the EXP30 Ultra SSD I/O drawer is at the following location:

http://pic.dhe.ibm.com/infocenter/powersys/v3r1m5/index.jsp?topic=/p7ham/p7ham_edr1_kickoff.htm

2.10.2 EXP24S SFF Gen2-bay drawer

The EXP24S SFF Gen2-bay drawer (FC 5887) is an expansion drawer that supports up to 24 hot-swap 2.5-inch SFF SAS HDDs on POWER6, POWER6+, POWER7, or POWER7+ server in 2U of 19-inch rack space.

The SFF bays of the EXP24S drawer differ from the SFF bays of the POWER7 or POWER7+ system unit or of the 12X PCIe I/O Drawers (FC 5802 or FC 5803). The EXP24S uses Gen2 or SFF-2 SAS drives that physically do not fit in the Gen1 or SFF-1 bays of the POWER7 or POWER7+ system unit, or of the 12X PCIe I/O Drawers.

The drawer can be attached to the Power 710 and Power 730 by either using the FC EJ0F storage backplane, providing an external SAS port, or using these PCIe SAS adapters or pair of adapters:

- PCIe LP 2-x4-Port SAS Adapter 3 Gb (FC 5278, CCIN 57B3)
- PCIe2 RAID SAS adapter Dual-Port 6 Gb (FC ESA2, CCIN 57B4)

In addition, the EXP24S drawer can also be connected to the integrated SAS controllers in the EXP30 Ultra SSD I/O drawer. The SAS controller and the EXP24S SAS ports are attached by using the appropriate SAS Y or X or EX cables.

Notes:

- A single FC 5887 drawer can be cabled to the CEC external SAS port when an FC EJ0F DASD backplane is part of the system. A 3 Gbps YI cable (FC 3686 or FC 3687) is used to connect an FC 5887 to the CEC external SAS port.
- A single FC 5887 is not allowed to attach to the CEC external SAS port when an FC EPCE processor (4-core) is ordered or installed on a single socket Power 710 system.

The SAS disk drives that are contained in the EXP24S drawer are controlled by one or two PCIe SAS adapters that are connected to the EXP24S through SAS cables. The SAS cable varies, depending on the adapter being used, the operating system being used, and the protection wanted.

In addition to the existing SAS disks options, IBM offers the following disk models:

- 900 GB 10K RPM SAS HDD in Gen-2 Carrier for AIX and Linux (FC 1752)
- 856 GB 10K RPM SAS HDD in Gen-2 Carrier for IBM i (FC 1738)

The EXP24S can be ordered in one of three possible manufacturing-configured mode settings (not customer set-up): 1, 2, or 4 sets of disk bays.

With IBM AIX, and Linux, the EXP24S drawer can be ordered with four sets of six bays (mode 4), two sets of 12 bays (mode 2), or one set of 24 bays (mode 1). With IBM i the EXP24S drawer can be ordered as one set of 24 bays (mode 1).

Notes:

- The modes for the EXP24S drawer are set by IBM manufacturing. There is no reset option after the drawer is shipped.
- If you order multiple EXP24S drawers, avoid mixing modes within that order. There is no externally visible indicator regarding the drawer's mode.
- Several EXP24S drawers cannot be cascaded on the external SAS connector. Only one FC 5887 is supported.
- The Power 710 and Power 730 support up to four EXP24S drawers.

Six SAS connectors are at the rear of the EXP24S drawer, to which the SAS adapters or controllers are attached. They are labeled T1, T2, and T3; there are two T1, two T2, and two T3 (Figure 2-32 on page 84):

- In mode 1, two or four of the six ports are used. Two T2 are used for a single SAS adapter, and two T2 and two T3 are used with a paired set of two adapters or dual adapters configuration.
- In mode 2 or mode 4, four ports will be used, two T2 and two T3, to access all SAS bays.

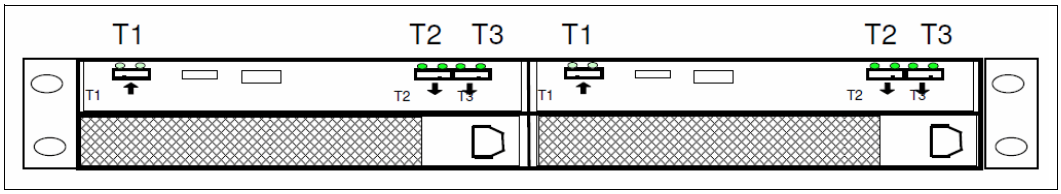


Figure 2-32 EXP24S SFF Gen2-bay drawer rear connectors

An EXP24S drawer in mode 4 can be attached to two or four SAS controllers and provide a high level of configuration flexibility. An EXP24S in mode 2 has similar flexibility. Up to 24 HDDs can be supported with any of the supported SAS adapters and controllers.

The EXP24S no-charge specify codes must be included with any EXP24S orders to indicate to IBM manufacturing the mode to which the drawer should be set and the adapter, controller, and cable configuration that will be used.

For details about the SAS cabling, see the serial-attached SCSI cable planning documentation:

http://pic.dhe.ibm.com/infocenter/powersys/v3r1m5/index.jsp?topic=/p7had/p7hadsascabling.htm

2.10.3 EXP12S SAS expansion drawer

The EXP12S (FC 5886) is an expansion drawer with twelve 3.5-inch form factor SAS bays. This drawer supports up to 12 hot-swap SAS HDDs or up to eight hot-swap SSDs. The EXP12S includes redundant AC power supplies and two power cords. Although the drawer is one set of 12 drives, which is run by one SAS controller or one pair of SAS controllers, it has two SAS attachment ports and two service managers for redundancy. The EXP12S occupies a 2U space in a 19-inch rack and the SAS controller can be a SAS PCIe adapter or pair of adapters.

The drawer can be attached to the Power 710 and Power 730 by either using the FC EJ0F storage backplane, providing an external SAS port, or using the following SAS adapters:

- PCIe LP Dual-x4-Port SAS adapter 3 Gb (FC 5278, CCIN 57B3)
- PCIe2 RAID SAS adapter Dual-Port 6 Gb (FC ESA2, CCIN 57B4)

A maximum number of eight EXP12S drawers can be attached to a Power 710 and Power 730 server.

Notes:

- An existing EXP12S SAS expansion drawer is supported, but no longer orderable.
- An existing EXP12S SAS expansion drawer is not supported on a 4-core Power 710 (FC EPCE).
- If the internal disk bay of the Power 710 or Power 730 server contains any SSD drives, an existing EXP12S SAS Expansion Drawer cannot be attached to the external SAS port on the Power 710 or Power 730. This rule applies even if the I/O drawer contains only SAS disk drives.

For details about SAS cabling, see the serial-attached SCSI cable planning documentation:

http://pic.dhe.ibm.com/infocenter/powersys/v3r1m5/index.jsp?topic=/p7had/p7hadsascabling.htm

2.10.4 IBM System Storage

The IBM System Storage Disk Systems products and offerings provide compelling storage solutions with superior value for all levels of business, from entry-level to high-end storage systems.

IBM System Storage N series

The IBM System Storage N series is a network-attached storage (NAS) solution. It offers the latest technology to help customers improve performance, virtualization manageability, and system efficiency at a reduced total cost of ownership. For more information about the IBM System Storage N series hardware and software, see the following location:

http://www.ibm.com/systems/storage/network

IBM Storwize V3700

IBM Storwize® V3700, the most recent addition to the IBM Storwize family of disk systems, delivers efficient, entry-level configurations, specifically designed to meet the needs of small and midsize businesses. Designed to provide organizations with the ability to consolidate and share data at an affordable price, Storwize V3700 offers advanced software capabilities that are usually found in more expensive systems. For more information, see the following site:

http://www.ibm.com/systems/storage/disk/storwize_v3700/index.html

IBM System Storage DS3500

IBM System Storage DS3500 combines best-of-its-kind development with leading 6 Gbps host interface and drive technology. With its simple, efficient, and flexible approach to storage, the DS3500 is a cost-effective, fully integrated complement to IBM System x® servers, IBM BladeCenter, and IBM Power Systems. By offering substantial improvements at a price that fits most budgets, the DS3500 delivers superior price for performance ratios, functionality, scalability, and ease of use for the entry-level storage user. For more information, see the following website:

http://www.ibm.com/systems/storage/disk/ds3500/index.html

IBM Storwize V7000 and Storwize V7000 Unified Disk Systems

IBM Storwize V7000 and IBM Storwize V7000 Unified are virtualized storage systems designed to consolidate workloads into a single storage system for simplicity of management, reduced cost, highly scalable capacity, performance, and high availability. They offer improved efficiency and flexibility through built-in solid-state drive (SSD) optimization, thin provisioning and nondisruptive migration of data from existing storage. They can also virtualize and reuse existing disk systems, offering a greater potential return on investment. Storwize V7000 and V7000 Unified now support integrated IBM Real-time Compression™, enabling storage of up to five times as much active primary data in the same physical space for extraordinary levels of efficiency.

The IBM Flex System™ V7000 Storage Node is also available as an integrated component of IBM Flex System and IBM PureFlex™ Systems and has been seamlessly integrated into the Flex System Manager and Chassis Map, delivering new data center efficiencies. For more information, see the following website:

http://www.ibm.com/systems/storage/disk/storwize_v7000/index.html

IBM XIV Storage System

IBM XIV® is a high-end disk storage system that helps thousands of enterprises meet the challenge of data growth with hot-spot-free performance and ease of use. Simple scaling, high service levels for dynamic, heterogeneous workloads, and tight integration with hypervisors and the OpenStack platform, enable optimal storage agility for cloud environments.

Optimized with inherent efficiencies that simplify storage, XIV delivers the benefits of IBM Smarter Storage for Smarter Computing, empowering organizations to take control of their storage and to extract more valuable insights from their data. XIV extends ease of use with integrated management for large and multiple site XIV deployments, reducing operational complexity and enhancing capacity planning. For more information, see the following website:

http://www.ibm.com/systems/storage/disk/xiv/index.html

IBM System Storage DS8000

The IBM System Storage DS8000® series is designed to manage a broad scope of storage workloads that exist in today's complex data center, doing it effectively and efficiently. The proven success of this flagship IBM disk system is a direct consequence of its extraordinary flexibility, reliability, and performance, but also of its capacity to satisfy the needs that are always changing. The latest evidence of DS8000 series value is the IBM System Storage DS8870 as the ideal storage platform for enterprise class environments. It provides unique performance, availability and scalability.

The DS8870 delivers the following benefits:

- Up to three times higher performance than DS8800.
- Improved security with full disk encryption (FDE) as standard on all systems.
- Optimized flash technology for dynamic performance and operational analytics.

Additionally, the DS8000 includes a range of features that automate performance optimization and application quality of service, and also provide the highest levels of reliability and system uptime. For more information, see the following website:

http://www.ibm.com/systems/storage/disk/ds8000/index.html

2.11 Hardware Management Console

The Hardware Management Console (HMC) is a dedicated appliance for configuring and managing system resources on IBM Power Systems servers that use IBM POWER5, POWER5+, POWER6, POWER6+ POWER7 and POWER7+ processors. The HMC provides basic virtualization management support for configuring logical partitions (LPARs) and dynamic resource allocation, including processor and memory settings for selected Power Systems servers. The HMC also supports advanced service functions, including guided repair and verify, concurrent firmware updates for managed systems, and error reporting on a continual basis, through IBM Electronic Service Agent™, for faster support.

The HMC management features help to improve server utilization, simplify systems management, and accelerate provisioning of server resources that use the PowerVM virtualization technology.

Requirements: When using the HMC with the Power 710 and Power 730 server, the HMC code must be running at V7R7.7.0 (SP1) level, or later.

The Power 710 and Power 730 platforms support two main service environments:

- Attachment to one or more HMCs is a supported option by the system

  This configuration is the common configuration for servers that support logical partitions with dedicated or virtual I/O. In this case, all servers have at least one logical partition.

- No HMC attachment

  Two service strategies are available for non HMC-attached systems:

  – Full system partition: A single partition owns all the server resources and only one operating system may be installed.

  – 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 includes some of the functions offered by the HMC.

Hardware support for customer-replaceable units (CRUs) is a standard inclusion, along with the HMC. In addition you have the option to upgrade this support level to IBM onsite support to be consistent with other Power Systems servers.

If you want to use an existing HMC to manage any POWER7+ processor-based server, the HMC must be model CR3 or later, rack-mounted HMC, or model C05 or later, deskside HMC.

HMC V7R7.7.0 is the last release to be supported on models 7310-C04, 7315-CR2, and 7310-CR2. Future HMC releases will not be supported on C04 or CR2 models.

When IBM Systems Director is used to manage an HMC or if the HMC manages more than 254 partitions, the HMC should have 3 GB of RAM minimum and be model CR3 or later, rack-mounted, or model C06 or later, deskside.

HMC code level

HMC V7R7.7.0 (SP1) contains the following new features:

- Support for managing IBM Power 710 and Power 730 systems
- Support for PowerVM functions such as new HMC GUI interface for VIOS install
- Improved transition from IVM to HMC management
- Ability to update a user password in Kerberos from the HMC for clients using remote HMC

HMC V7R7.7.0 (SP1) supports up to 48 servers (non Power 590, Power 595, and Power 795 models) or 32 IBM Power 590, Power 595, and Power 795 servers. A maximum of 2000 LPARs are supported when you use a HMC V7R7.6.0 code at a minimum level and the HMC is model 7042-CR6 or later.

If you attach an existing HMC to a new server such as the Power 710 and Power 730 or add functions to an existing server that requires a firmware update, the HMC machine code might need to be updated. Upgrade the support level of the HMC to be consistent with the support that is provided on the servers to which it is attached. In a dual HMC configuration, both systems must be at the same version and release of the HMC.

To determine the HMC machine code level that is required for the firmware level on any server, go to the following website to access Fix Central and the Fix Level Recommendation Tool (FLRT) on or after the planned availability date for this product. FLRT will identify the correct HMC machine code for the selected system firmware level.

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

With HMC code V7R7.7.0 (SP1), the HMC supports Mozilla Firefox 7 through 10 and Microsoft Internet Explorer 7 through 9.

HMC RAID 1 support

HMCs now offer a high-availability feature. Starting from HMC 7042-CR7 RAID 1 protection will be enabled by default. This feature enables data redundancy by using two physical disk drives.

RAID 1 is also offered on both the 7042-CR6 and the 7042-CR7 (if the feature was removed from the initial order) as an MES upgrade option.

Blade management

The HMC gives systems administrators a tool for planning, virtualizing, deploying, and managing IBM Power System servers.

With the introduction of HMC V7R760, the HMC can now manage IBM BladeCenter Power Blade servers. This management includes support for dual VIOS, live partition mobility between blades and rack servers, and management of both blades and rack servers from a single management console.

Comparison of 7042-CR6 and 7042-CR7 HMC models

The 7042-CR6 was withdrawn from marketing in December 2012. For your reference, Table 2-25 compares features of the 7042-CR6 and the 7042-CR7 HMC models.

Table 2-25 Comparison for 7042-CR6 and 7042-CR7

Feature	CR6	CR7
IBM System x model	x3550 M3	x3550 M4
HMC model	7042-CR6	7042-CR7
Processor	Westmere-EP	Intel Xeon E5
Memory	4 GB	4 GB
DASD	500 GB	500 GB
RAID 1	Optional	Default
Multitech internal modem	Default	Optional
USB ports	Two front, four back, one internal	Two front, four back, one internal
Integrated network	Two on main bus plus two on expansion slot	Four 1 Gb Ethernet
I/O slots	1 PCI Express 2.0 slot	1 PCI Express 3.0 slot

2.11.1 HMC connectivity to the POWER7+ processor-based systems

POWER7+ processor technology-based servers and their predecessor systems that are managed by an HMC require Ethernet connectivity between the HMC and the server's service processor. In addition, if dynamic LPAR, Live Partition Mobility, or PowerVM Active Memory Sharing operations are required on the managed partitions, Ethernet connectivity is needed between these partitions and the HMC. A minimum of two Ethernet ports are needed on the HMC to provide such connectivity.

For any logical partition in a server, you may use a Shared Ethernet Adapter that is configured through a Virtual I/O Server. Therefore, a partition does not require its own physical adapter to communicate with an HMC.

For the HMC to communicate properly with the managed server, eth0 port of the HMC must be connected to either the HMC1 or HMC2 ports of the managed server, although other network configurations are possible. You can attach a second HMC to HMC2 port of the server for redundancy (or vice versa). These ports must be addressed by two separate subnets. Figure 2-33 shows a simple network configuration to enable the connection from the HMC to the server and to enable dynamic LPAR operations. For more details about HMC and the possible network connections, see *IBM Power Systems HMC Implementation and Usage Guide*, SG24-7491 (previous edition was named *Hardware Management Console V7 Handbook*, SG24-7491).

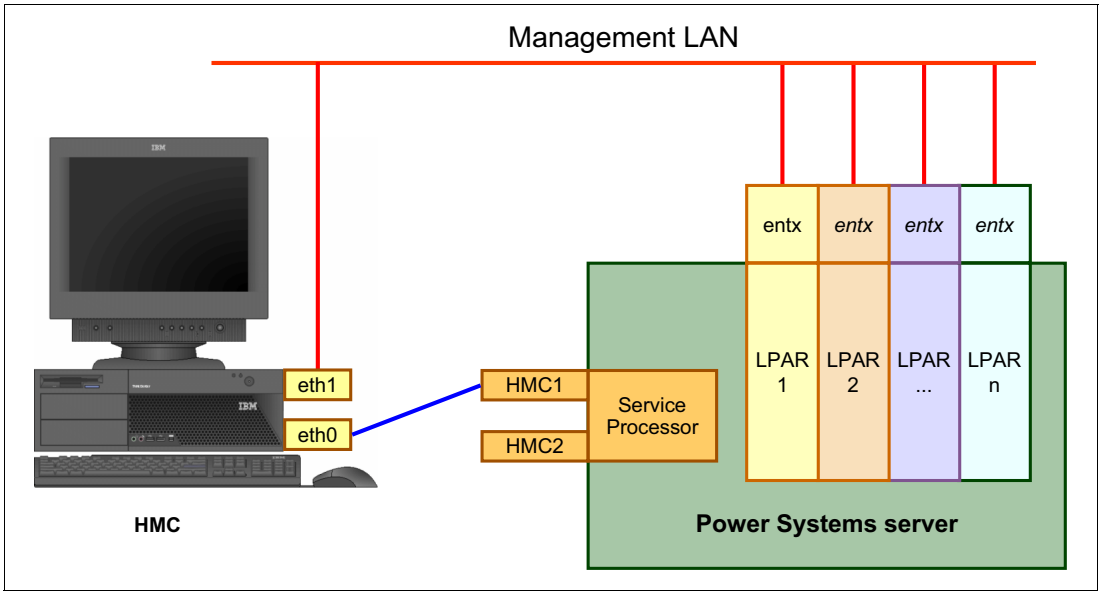


Figure 2-33 HMC to service processor and LPARs network connection

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 that the managed server is powered on. In this case, the flexible service processors (FSPs) are allocated an IP address from a set of address ranges that are predefined in the HMC software. These predefined ranges are identical for version V7R7.1.0 of the HMC code and for previous versions.

If the service processor of the managed server does not receive a DHCP reply before time out, predefined IP addresses are 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.

Notes: The service processor is used to monitor and manage the system hardware resources and devices. The two service processor HMC ports run at a speed of 100 Mbps.

- Both HMC ports are visible only to the service processor and can be used to attach the server to an HMC, or to access the ASMI options from a client web browser by using the HTTP server that is integrated into the service processor internal operating system.
- When no IP address is set, by default, the configurations are as follows:
  - 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.

For more information about the service processor, see “Service processor” on page 162.

2.11.2 High availability HMC configuration

The HMC is an important hardware component. When in operation, Power Systems servers and their hosted partitions can continue to operate when no HMC is available. However, in such conditions, certain operations cannot be performed, such as a dynamic LPAR reconfiguration, a partition migration using PowerVM Live Partition Mobility, or the creation of a new partition. You might therefore decide to install two HMCs in a redundant configuration so that one HMC is always operational, even when performing maintenance of the other one, for example.

If redundant HMC functionality is what you want, a server can be attached to two independent HMCs to address availability requirements. Both HMCs must have the same level of Hardware Management Console Licensed Machine Code Version 7 and installed fixes to manage POWER7+ processor-based servers, or an environment with a mixture of POWER5, POWER5+, POWER6, POWER6+, POWER7, and POWER7+ processor-based servers. The HMCs provide a locking mechanism so that only one HMC at a time has write access to the service processor. Both HMCs should be available on a public subnet to allow full synchronization of functionality. Depending on your environment, you have multiple options to configure the network.

Figure 2-34 shows one possible highly available HMC configuration that is managing two servers. Each HMC is connected to one FSP port of all managed servers.

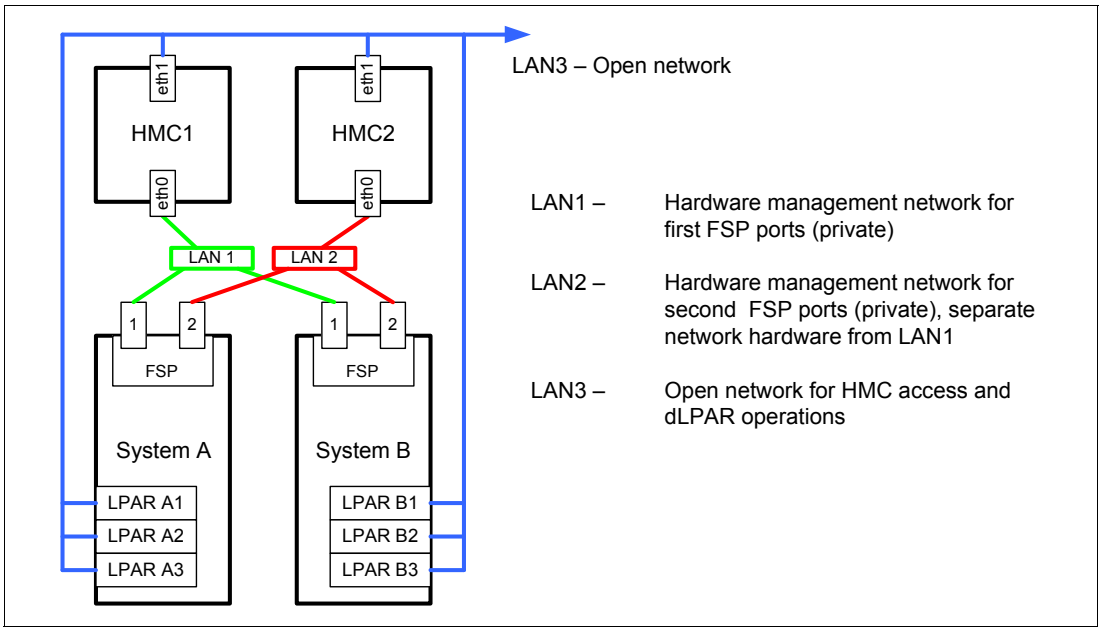


Figure 2-34 Highly available HMC and network architecture

For simplicity, only hardware management networks (LAN1 and LAN2) are highly available (Figure 2-34). However, the management network (LAN3) can be made highly available by using a similar concept and adding more Ethernet adapters to the LPARs and HMCs.

Both HMCs must be on a separate virtual local area network (VLAN) to protect from any network contention. Each HMC can be a DHCP server for its VLAN.

For details about redundant HMC, see *IBM Power Systems HMC Implementation and Usage Guide*, SG24-7491 (previous edition was named *Hardware Management Console V7 Handbook*, SG24-7491).

If you want to migrate an LPAR from a POWER6 processor-based server onto a POWER7+ processor-based server by using PowerVM Live Partition Mobility, consider how the source server is managed. If the source server is managed by one HMC and the destination server is managed by another HMC, ensure that the HMC that is managing the POWER6 processor-based server is at a minimum level of V7R7.3.5 or later, and that the HMC that is managing the POWER7+ processor-based server is at minimum level of V7R7.6.0 or later.

2.12 Operating system support

The Power 710 and Power 730 servers support the following operating systems:

- AIX
- IBM i
- Linux

In addition, the Virtual I/O Server can be installed in special partitions that provide support to the other operating systems for using features such as virtualized I/O devices, PowerVM Live Partition Mobility, or PowerVM Active Memory Sharing.

For details about the software available on IBM Power Systems, visit the IBM Power Systems Software™ website:

http://www.ibm.com/systems/power/software/index.html

2.12.1 IBM AIX operating system

IBM periodically releases maintenance packages (service packs or technology levels) for the AIX operating system. Information about these packages, downloading, and obtaining the CD-ROM is on the Fix Central website. The Fix Central website also provides information about how to obtain the fixes that are included on CD-ROM.

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

The Service Update Management Assistant (SUMA), which can help you to automate the task of checking and downloading operating systems, is part of the base operating system. For more information about the **suma** command, go to the following website:

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

IBM AIX Version 5.3

At the time of writing, AIX Version 5.3 is not supported with the Power 710 and Power 730.

Statement of Direction (SoD): IBM intends to provide to those clients with AIX 5.3 Technology Level 12 (and the associated service extension offering) the ability to run that environment on Power 710 and Power 730.

IBM AIX Version 6.1

The following minimum levels of AIX Version 6.1 support the Power 710 and Power 730:

- AIX V6.1 with the 6100-08 Technology Level and Service Pack 2, or later
- AIX V6.1 with the 6100-07 Technology Level and Service pack 7, or later (planned availability March 29, 2013)
- AIX V6.1 with the 6100-06 Technology Level and Service pack 11, or later (planned availability March 29, 2013)

A partition that uses AIX 6.1 can run in POWER6, POWER6+, or POWER7 mode. The best approach is to run the partition in POWER7 mode to allow exploitation of new hardware capabilities such as SMT4 and Active Memory Expansion.

IBM AIX Version 7.1

The following minimum level of AIX Version 7.1 supports the Power 710 and Power 730:

- AIX V7.1 with the 7100-02 Technology Level and Service Pack 2, or later

Statement of Direction (SoD): IBM intends to provide to those clients with AIX 7.1 Technology Level 0 or Technology Level 1 the ability to run that environment on Power 710 and Power 730.

A partition that uses AIX 7.1 can run in POWER6, POWER6+, or POWER7 mode. The best approach is to run the partition in POWER7 mode to allow exploitation of new hardware capabilities such as SMT4 and Active Memory Expansion.

2.12.2 IBM i operating system

The IBM i operating system is supported on the Power 710 and Power 730 with the following minimum required levels:

- IBM i 7.1, or later
- IBM i 6.1 with machine code 6.1.1, or later
  - Requires all I/O to be virtual
  - Cannot be ordered as the primary operating system with FC 2145 and FC 0566

IBM periodically releases maintenance packages (service packs or technology levels) for the IBM i operating system. Information about these packages, downloading, and obtaining the CD-ROM is on the Fix Central website:

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

Visit the IBM Prerequisite website for compatibility information for hardware features and the corresponding AIX and IBM i Technology Levels.

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

2.12.3 Linux operating system

Linux is an open source operating system that runs on numerous platforms, from embedded systems to mainframe computers. It provides an implementation like UNIX across many computer architectures.

The supported versions of Linux on the Power 710 and Power 730 servers are as follows:

- SUSE Linux Enterprise Server 11 Service Pack 2, or later, with current maintenance updates available from Novell to enable all planned functionality
- For Red Hat Enterprise Linux (RHEL), consult the following Statements of Direction:
  - RHEL 6.4 support for Power 710 and Power 730

    IBM intends to continue to work with Red Hat to provide support for Power 710 and Power 730 with an upcoming Red Hat Enterprise Linux 6 release. For additional questions about the availability of this release and supported hardware servers, consult the Red Hat Hardware Catalog at:

    https://hardware.redhat.com
  - RHEL 6 preinstall feature for Power 710 and Power 730

    IBM intends to provide support for preinstall of an upcoming Red Hat Enterprise Linux 6 release on the Power 710 and Power 730 systems.

If you want to configure Linux partitions in virtualized Power Systems, be aware of the following conditions:

- 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 the POWER7+ processor-based servers, or from other Linux distributors.

For information about features and external devices that are supported by Linux, see the following site:

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

Be sure to update your systems with the latest Linux for Power service and productivity tools:

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

See information about SUSE Linux Enterprise Server:

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

See information about Red Hat Enterprise Linux Advanced Server:

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

2.12.4 Virtual I/O Server

The minimum required level of Virtual I/O Server for both the Power 710 and Power 730 is VIOS 2.2.2.2.

Statement of Direction (SoD): IBM intends to provide to those clients with VIOS 2.2.1 the ability to run that environment on the Power 710 and Power 730.

IBM regularly updates the Virtual I/O Server code. To find information about the latest updates, visit the Fix Central website:

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

2.12.5 Java versions that are supported

Unique considerations exist for running Java 1.4.2 on POWER7 or POWER7+ servers. For best use of the performance capabilities and most recent improvements of POWER7 technology, upgrade Java-based applications to Java 7, Java 6, or Java 5 when possible. See the AIX download and service information page:

http://www.ibm.com/developerworks/java/jdk/aix/service.html

2.12.6 Boosting performance and productivity with IBM compilers

IBM XL C, XL C/C++, and XL Fortran compilers for AIX and for Linux use the latest POWER7+ processor architecture. With each release, these compilers continue to help improve application performance and capability, exploiting architectural enhancements that are made available through the advancement of the POWER technology.

IBM compilers are designed to optimize and tune your applications for execution on IBM POWER platforms to help you unleash the full power of your IT investment, to create and maintain critical business and scientific applications, to maximize application performance, and to improve developer productivity.

The performance gain from years of compiler optimization experience is seen in the continuous release of compiler improvements that support the POWER4 processors, through to POWER4+, POWER5, POWER5+, POWER6, and POWER7 processors, and now including the POWER7+ processors. With the support of the latest POWER7+ processor chip, IBM advances a more than a 20-year investment in the XL compilers for POWER series and IBM PowerPC® series architectures.

The XL C, XL C/C++, and XL Fortran features that are introduced to use the latest POWER7+ processor include the following items:

- Vector unit and vector scalar extension (VSX) instruction set to efficiently manipulate vector operations in your application
- Vector functions within the Mathematical Acceleration Subsystem (MASS) libraries for improved application performance
- Built-in functions or intrinsics and directives for direct control of POWER instructions at the application level
- Architecture and tune compiler options to optimize and tune your applications

COBOL for AIX enables you to selectively target code generation of your programs to either exploit POWER7+ systems architecture or to be balanced among all supported POWER systems. The performance of COBOL for AIX applications is improved by means of an enhanced back-end optimizer. With the back-end optimizer, a component common also to the IBM XL compilers, your applications can use the most recent industry-leading optimization technology.

The performance of PL/I for AIX applications is improved through both front-end changes and back-end optimizer enhancements. With the back-end optimizer, a component common also to the IBM XL compilers, your applications can use the most recent industry-leading optimization technology. For PL/I, it produces code that is intended to perform well across all hardware levels, including POWER7+ of AIX.

IBM Rational® Development Studio for IBM i 7.1 provides programming languages for creating modern business applications:

- ILE RPG
- ILE COBOL
- C and C++ compilers
- Heritage RPG and COBOL compilers

The latest release includes performance improvements and XML processing enhancements for ILE RPG and ILE COBOL, improved COBOL portability with a COMP-5 data type, and easier Unicode migration with relaxed USC2 rules in ILE RPG. Rational also released a product named Rational Open Access: RPG Edition. This product opens the ILE RPG file I/O processing, enabling partners, tool providers, and users to write custom I/O handlers that can access other devices like databases, services, and web user interfaces.

IBM Rational Developer for Power Systems Software provides a rich set of integrated development tools that support the XL C/C++ for AIX compiler, the XL C for AIX compiler, and the COBOL for AIX compiler. Rational Developer for Power Systems Software offers capabilities of file management, searching, editing, analysis, build, and debug, all integrated into an Eclipse workbench. XL C/C++, XL C, and COBOL for AIX developers can boost productivity by moving from older, text-based, command-line development tools to a rich set of integrated development tools.

The IBM Rational Power Appliance solution provides a workload-optimized system and integrated development environment for AIX development on IBM Power Systems. IBM Rational Power Appliance includes a Power Express server, preinstalled with a comprehensive set of Rational development software along with the AIX operating system. The Rational development software includes support for Collaborative Application Lifecycle Management (C/ALM) through IBM Rational Team Concert™, a set of software development tools from Rational Developer for Power Systems Software, and a choice between the XL C/C++ for AIX or COBOL for AIX compilers.

2.13 Energy management

The Power 710 and 730 servers are designed with features to help clients become more energy efficient. The IBM Systems Director Active Energy Manager uses EnergyScale technology, enabling advanced energy management features to dramatically and dynamically conserve power and further improve energy efficiency. Intelligent Energy optimization capabilities enable the POWER7+ processor to operate at a higher frequency for increased performance and performance per watt, or dramatically reduce frequency to save energy.

Certain configurations of the Power 730 server are ENERY STAR qualified. For details see the following website:

http://www.ibm.com/systems/hardware/energy_star/power.html

2.13.1 IBM EnergyScale technology

IBM EnergyScale technology provides functions to help the user understand and dynamically optimize processor performance versus processor energy consumption, and system workload, to control IBM Power Systems power and cooling usage.

On POWER7 or POWER7+ processor-based systems, the thermal power management device (TPMD) card is responsible for collecting the data from all system components, changing operational parameters in components, and interacting with the IBM Systems Director Active Energy Manager (an IBM Systems Director plug-in) for energy management and control.

IBM EnergyScale uses power and thermal information that is collected from the system to implement policies that can lead to better performance or better energy utilization. IBM EnergyScale has the following features:

- Power trending

  EnergyScale provides continuous collection of real-time server energy consumption. It enables administrators to predict power consumption across their infrastructure and to react to business and processing needs. For example, administrators can use such information to predict data center energy consumption at various times of the day, week, or month.

- Thermal reporting

  IBM Director Active Energy Manager can display measured ambient temperature and calculated exhaust heat index temperature. This information can help identify data center hot spots that need attention. See Figure 2-35 on page 97 for an example.

- Power saver mode

  Power saver mode lowers the processor frequency and voltage on a fixed amount, reducing the energy consumption of the system while still delivering predictable performance. This percentage is predetermined to be within a safe operating limit and is not user configurable. The server is designed for a fixed frequency drop of almost 50% down from nominal frequency (the actual value depends on the server type and configuration).

  Power saver mode is not supported during booting or rebooting, although it is a persistent condition that is sustained after the boot when the system starts executing instructions.

- Dynamic power saver mode

  Dynamic power saver mode varies processor frequency and voltage based on the utilization of the POWER7 or POWER7+ processors. Processor frequency and utilization

are inversely proportional for most workloads, implying that as the frequency of a processor increases, its utilization decreases, given a constant workload. Dynamic power saver mode takes advantage of this relationship to detect opportunities to save power, based on measured real-time system utilization.

When a system is idle, the system firmware lowers the frequency and voltage to power energy saver mode values. When fully utilized, the maximum frequency varies, depending on whether the user favors power savings or system performance. If an administrator prefers energy savings and a system is fully utilized, the system is designed to reduce the maximum frequency to about 95% of nominal values. If performance is favored over energy consumption, the maximum frequency can be increased to up to 111.6% of nominal frequency for extra performance.

Table 2-26 shows the maximum frequency increases of the various processor options.

Table 2-26 Maximum frequency increase values for Power 710 and Power 730

Processor module option	Power 710	Power 730
3.6 GHz 4-core (FC EPCE)	11.6%	
3.6 GHz 8-core (FC EPCH)		11.6%
4.2 GHz 6-core (FC EPCG)	5.9%	5.9%
4.2 GHz 8-core (FC EPCJ)	7.3%	7.3%
4.3 GHz 4-core (FC EPCF)		5.8%

Dynamic power saver mode is mutually exclusive with power saver mode. Only one of these modes can be enabled at a given time.

Figure 2-35 provides a view shown by the Active Energy Manager that shows the dynamic CPU frequency change in a system using the Dynamic power saver mode.

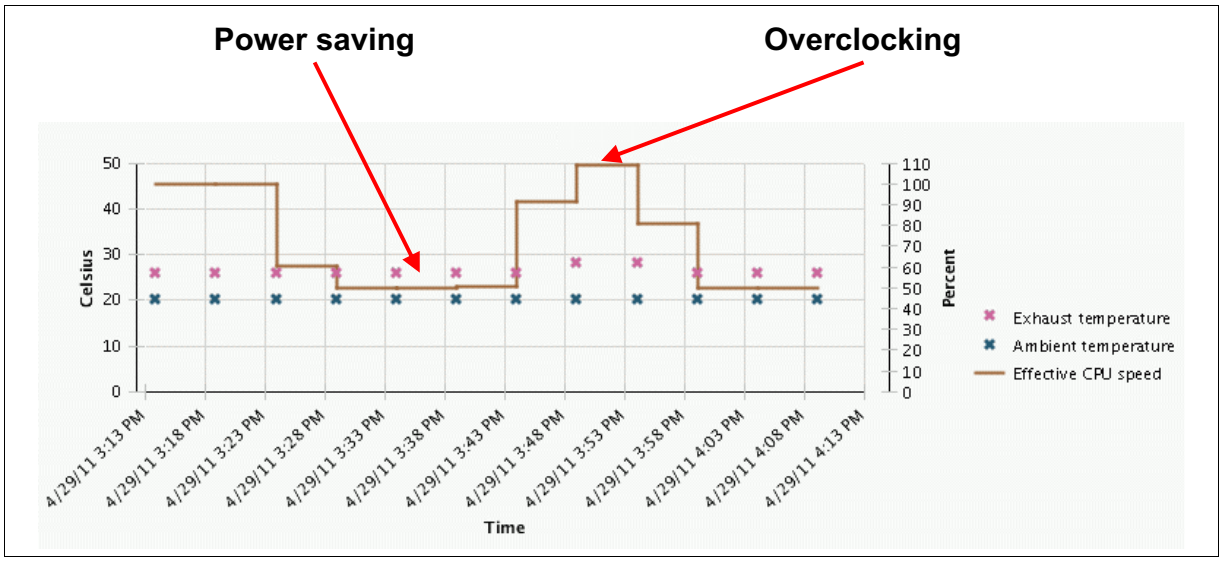


Figure 2-35 Example of a system using Dynamic Power saver mode

- Power capping

  Power capping enforces a user-specified limit on power usage. Power capping is not a power-saving mechanism. It enforces power caps by throttling the processors in the system, degrading performance significantly. The idea of a power cap is to set a limit that

must never be reached but that frees extra power that was never used in the data center. The *margined* power is this amount of extra power that is allocated to a server during its installation in a data center. It is based on the server environmental specifications that usually are never reached because server specifications are always based on maximum configurations and worst-case scenarios. The user must set and enable an energy cap from the IBM Director Active Energy Manager user interface.

- Soft power capping

  There are two power ranges into which the power cap can be set: power capping, as described previously, and soft power capping. Soft power capping extends the allowed energy capping range further, beyond a region that can be guaranteed in all configurations and conditions. If the energy management goal is to meet a particular consumption limit, then soft power capping is the mechanism to use.

- Processor core nap mode

  IBM POWER7 and POWER7+ processor uses a low-power mode called *nap* that stops processor execution when there is no work to do on that processor core. The latency of exiting nap mode is small, typically not generating any impact on applications running. Therefore, the IBM POWER Hypervisor™ can use nap mode as a general-purpose idle state. When the operating system detects that a processor thread is idle, it yields control of a hardware thread to the POWER Hypervisor. The POWER Hypervisor immediately puts the thread into nap mode. Nap mode allows the hardware to turn the clock off on most of the circuits in the processor core. Reducing active energy consumption by turning off the clocks allows the temperature to fall, which further reduces leakage (static) power of the circuits causing a cumulative effect. Nap mode saves 10 - 15% of power consumption in the processor core.

- Processor core sleep mode

  To be able to save even more energy, the POWER7+ processor has an even lower power mode referred to as *sleep*. Before a core and its associated private L2 cache enter sleep mode, the cache is flushed, transition lookaside buffers (TLB) are invalidated, and the hardware clock is turned off in the core and in the cache. Voltage is reduced to minimize leakage current. Processor cores that are inactive in the system (such as capacity on demand, CoD, processor cores) are kept in sleep mode. Sleep mode saves about 80% power consumption in the processor core and its associated private L2 cache.

- Processor chip winkle mode

  The most amount of energy can be saved when a whole POWER7+ chiplet enters the *winkle* mode. In this mode the entire chiplet is turned off including the L3 cache. This way can save more than 95% power consumption.

- Fan control and altitude input

  System firmware dynamically adjusts fan speed based on energy consumption, altitude, ambient temperature, and energy savings modes. Power Systems are designed to operate in worst-case environments, in hot ambient temperatures, at high altitudes, and with high-power components. In a typical case, one or more of these constraints are not valid. When no power savings setting is enabled, fan speed is based on ambient temperature and assumes a high-altitude environment. When a power savings setting is enforced (either Power Energy Saver Mode or Dynamic Power Saver Mode), fan speed will vary based on power consumption, ambient temperature, and altitude available. System altitude can be set in IBM Director Active Energy Manager. If no altitude is set, the system will assume a default value of 350 meters above sea level.

  The Power 710 and the Power 730 comply to the ASHRAE Class A3 standard and can support up to 35 degrees Celsius and 1825 meter at the rated performance. However, they could operate in a degraded performance above 35 degrees Celsius up to 40 degrees Celsius, or higher altitudes.

- Processor folding

  Processor folding is a consolidation technique that dynamically adjusts, over the short term, the number of processors available for dispatch to match the number of processors demanded by the workload. As the workload increases, the number of processors made available increases. As the workload decreases, the number of processors that are made available decreases. Processor folding increases energy savings during periods of low to moderate workload because unavailable processors remain in low-power idle states (nap or sleep) longer.

- EnergyScale for I/O

  IBM POWER7 and POWER7+ processor-based systems automatically power off hot pluggable PCI adapter slots that are empty or not being used. 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 POWER7 and POWER7+ processor-based servers and the expansion units that they support.

- Server power down

  If overall data center processor utilization is low, workloads can be consolidated on fewer numbers of servers so that some servers can be turned off completely. Consolidation makes sense when there will be long periods of low utilization, such as weekends. Active Energy Manager (AEM) provides information, such as the power that will be saved and the time needed to bring a server back online, that can be used to help make the decision to consolidate and power off. As with many of the features that are available in IBM Systems Director and Active Energy Manager, this function is scriptable and can be automated.

- Partition power management

  Available with Active Energy Manager 4.3.1 or later, and POWER7 systems with the 730 firmware release or later, is the capability to set a power savings mode for partitions or the system processor pool. As in the system-level power savings modes, the per-partition power savings modes can be used to achieve a balance between the power consumption and the performance of a partition. Only partitions that have dedicated processing units can have a unique power savings setting. Partitions that run in shared processing mode have a common power savings setting, which is that of the system processor pool. The reason is because processing unit fractions cannot be power-managed.

  As in the case of system-level power savings, two Dynamic Power Saver options are offered:

  - Favor partition performance
  - Favor partition power savings

  This setting must configured from Active Energy Manager. When dynamic power saver is enabled in either mode, system firmware continuously monitors the performance and utilization of each of the computer's POWER7 or POWER7+ processor cores that belong to the partition. Based on this utilization and performance data, the firmware dynamically adjusts the processor frequency and voltage, reacting within milliseconds to adjust workload performance and also deliver power savings when the partition is underused.

  In addition to the two dynamic power saver options, the customer can select to have no power savings on a given partition. This option leaves the processor cores that are assigned to the partition running at their nominal frequencies and voltages.

  A power savings mode, referred to as *inherit host setting*, is available and is applicable only to partitions. When configured to use this setting, a partition adopts the power savings mode of its hosting server. By default, all partitions with dedicated processing units, and the system processor pool, are set to the inherit host setting.

On POWER7 and POWER7+ processor-based systems, several EnergyScale technologies are imbedded in the hardware and do not require an operating system or external management component. More advanced functionality requires Active Energy Manager (AEM) and IBM Systems Director.

Table 2-27 lists all features that are supported, showing all cases in which AEM is not required, and also detailing the features that can be activated by traditional user interfaces (for example, ASMI and HMC).

Table 2-27 AEM support

Feature	AEM required	ASMI	HMC
Power Trending	Yes	No	No
Thermal Reporting	Yes	No	No
Static Power Saver	No	Yes	Yes
Dynamic Power Saver	Yes	No	No
Power Capping	Yes	No	No
Energy-optimized Fans	No	-	-
Processor Core Nap	No	-	-
Processor Core Sleep	No	-	-
Processor Winkle mode	No	-	-
Processor Folding	No	-	-
EnergyScale for I/O	No	-	-
Server Power Down	Yes	-	-
Partition Power Management	Yes	-	-

The Power 710 and Power 730 systems implement all EnergyScale capabilities that are listed in 2.13.1, “IBM EnergyScale technology” on page 96.

2.13.2 Thermal power management device card

The Thermal power management device (TPMD) card is a separate micro controller installed on some POWER6 processor-based systems, and on all POWER7 and POWER7+ processor-based systems. It runs real-time firmware whose sole purpose is to manage system energy.

The TPMD card monitors the processor modules, memory, environmental temperature, and fan speed. Based on this information, it can act upon the system to maintain optimal power and energy conditions (for example, increase the fan speed to react to a temperature change). It also interacts with the IBM Systems Director Active Energy Manager to report power and thermal information and to receive input from AEM on policies to be set. The TPMD is part of the EnergyScale infrastructure.

2.13.3 Energy consumption estimation

Often, for Power Systems, various energy-related values are important:

- Maximum power consumption and power source loading values

  These values are important for site planning and are in the hardware information center:

  http://pic.dhe.ibm.com/infocenter/powersys/v3r1m5/index.jsp

  Search for type and model number and `server specifications`. For example, for the Power 730 system search for `8231-E2D server specifications`.

- An estimation of the energy consumption for a certain configuration

  The calculation of the energy consumption for a certain configuration could be done in the IBM Systems Energy Estimator:

  http://www-912.ibm.com/see/EnergyEstimator/

  In that tool select the type and model for the desired system, enter some details of the configuration and a desired CPU utilization. As a result the tool shows the estimated energy consumption and the waste heat at the desired utilization and also at full utilization.

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 offer key capabilities that can help you consolidate and simplify your IT environment:

- Improve server utilization and sharing 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, so you can 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:

- POWER Hypervisor
- POWER processor modes
- Active Memory Expansion
- PowerVM
- System Planning Tool
- New PowerVM version 2.2.2 features

© Copyright IBM Corp. 2013. All rights reserved.

3.1 POWER Hypervisor

Combined with features in the POWER7+ processors, the IBM POWER Hypervisor delivers functions that enable other system technologies, including logical partitioning technology, virtualized processors, IEEE VLAN-compatible virtual switch, virtual SCSI adapters, virtual Fibre Channel 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 "Processing mode" on page 116).
- 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 performs a reset or 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 also when not connected to the managed console. It requires memory to support the resource assignment to the logical partitions on the server. The amount of memory that is required by the POWER Hypervisor firmware varies according to several factors:

- Number of logical partitions
- Number of physical and virtual I/O devices used by the logical partitions
- Maximum memory values specified in the logical partition profiles

The minimum amount of physical memory that is required to create a partition will be the size of the system's logical memory block (LMB). The default LMB size varies according to the amount of memory that is configured in the CEC (Table 3-1).

Table 3-1 Configured CEC memory-to-default logical memory block size

Configurable CEC memory	Default logical memory block
Up to 32 GB	128 MB
Greater than 32 GB	256 MB

In most cases, however, the actual minimum requirements and recommendations of the supported operating systems are greater than 256 MB. Physical memory is assigned to partitions in increments of LMB.

The POWER Hypervisor provides the following types of virtual I/O adapters:

- Virtual SCSI
- Virtual Ethernet
- Virtual Fibre Channel
- Virtual (TTY) console

Virtual SCSI

The POWER Hypervisor provides a virtual SCSI mechanism for the virtualization of storage devices. The storage virtualization is accomplished by using two paired adapters:

- A virtual SCSI server adapter
- A virtual SCSI client adapter

A Virtual I/O Server partition or an IBM i partition can define virtual SCSI server adapters. Other partitions are *client* partitions. The Virtual I/O Server partition is a special logical partition, as described in 3.4.4, “Virtual I/O Server” on page 122. The Virtual I/O Server software is included on all PowerVM editions. When using the PowerVM Standard Edition and PowerVM Enterprise Edition, dual Virtual I/O Servers can be deployed to provide maximum availability for client partitions when performing Virtual I/O Server maintenance.

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 up to 20 Gbps, depending on the maximum transmission unit (MTU) size, type of communication and CPU entitlement. Virtual Ethernet support began with IBM AIX Version 5.3, Red Hat Enterprise Linux 4, and SUSE Linux Enterprise Server, 9, and it is supported on all later versions. (For more information, see 3.4.10, “Operating system support for PowerVM” on page 134). 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 65,408 bytes. Therefore, the maximum MTU for the corresponding interface can be up to 65,394 (or 65,390 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 can support 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, also known as Shared Ethernet Adapter. See 3.4.4, “Virtual I/O Server” on page 122, for details about shared Ethernet.

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

Virtual Fibre Channel

A virtual Fibre Channel adapter is a virtual adapter that provides client logical partitions with a Fibre Channel connection to a storage area network through the Virtual I/O Server logical partition. The Virtual I/O Server logical partition provides the connection between the virtual Fibre Channel adapters on the Virtual I/O Server logical partition and the physical Fibre Channel adapters on the managed system. Figure 3-1 depicts the connections between the client partition virtual Fibre Channel adapters and the external storage. For additional information, see 3.4.10, “Operating system support for PowerVM” on page 134.

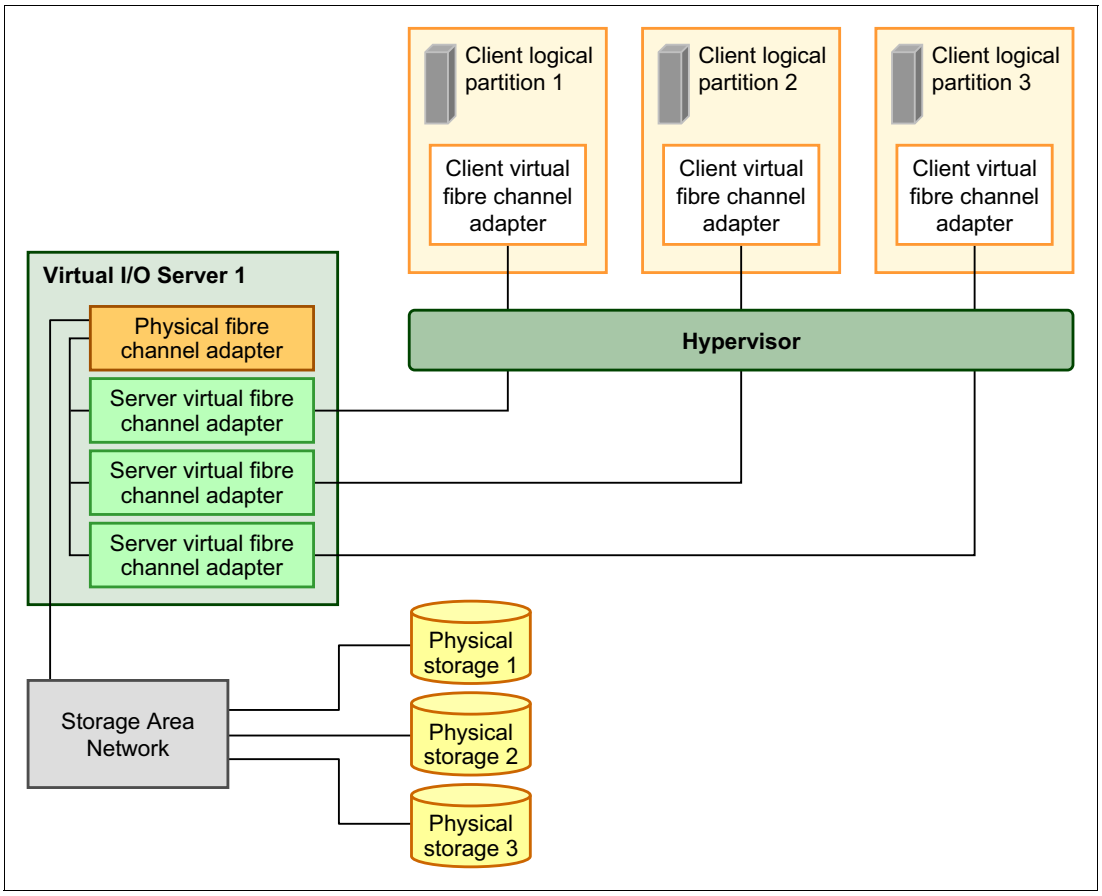


Figure 3-1 Connectivity between virtual Fibre Channels adapters and external SAN devices

Virtual (TTY) console

Each partition must have access to a system console. Tasks such as operating system installation, network setup, and various problem analysis activities require a dedicated system console. The POWER Hypervisor provides the virtual console by 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 POWER processor modes

Although, strictly speaking, not a virtualization feature, the POWER modes are described here because they affect various virtualization features.

On Power System servers, partitions can be configured to run in several modes, including the following modes:

- POWER6 compatibility mode

  This execution mode is compatible with Version 2.05 of the Power Instruction Set Architecture (ISA). For more information, visit the following address:

  http://power.org/wp-content/uploads/2012/07/PowerISA_V2.05.pdf

- POWER6+ compatibility mode

  This mode is similar to POWER6, with eight more storage protection keys.

- POWER7 mode

  This is the native mode for POWER7+ and POWER7 processors, implementing the v2.06 of the Power Instruction Set Architecture. For more information, visit the following address:

  http://power.org/wp-content/uploads/2012/07/PowerISA_V2.06B_V2_PUBLIC.pdf

The selection of the mode is made on a per-partition basis, from the managed console, by editing the partition profile (Figure 3-2).

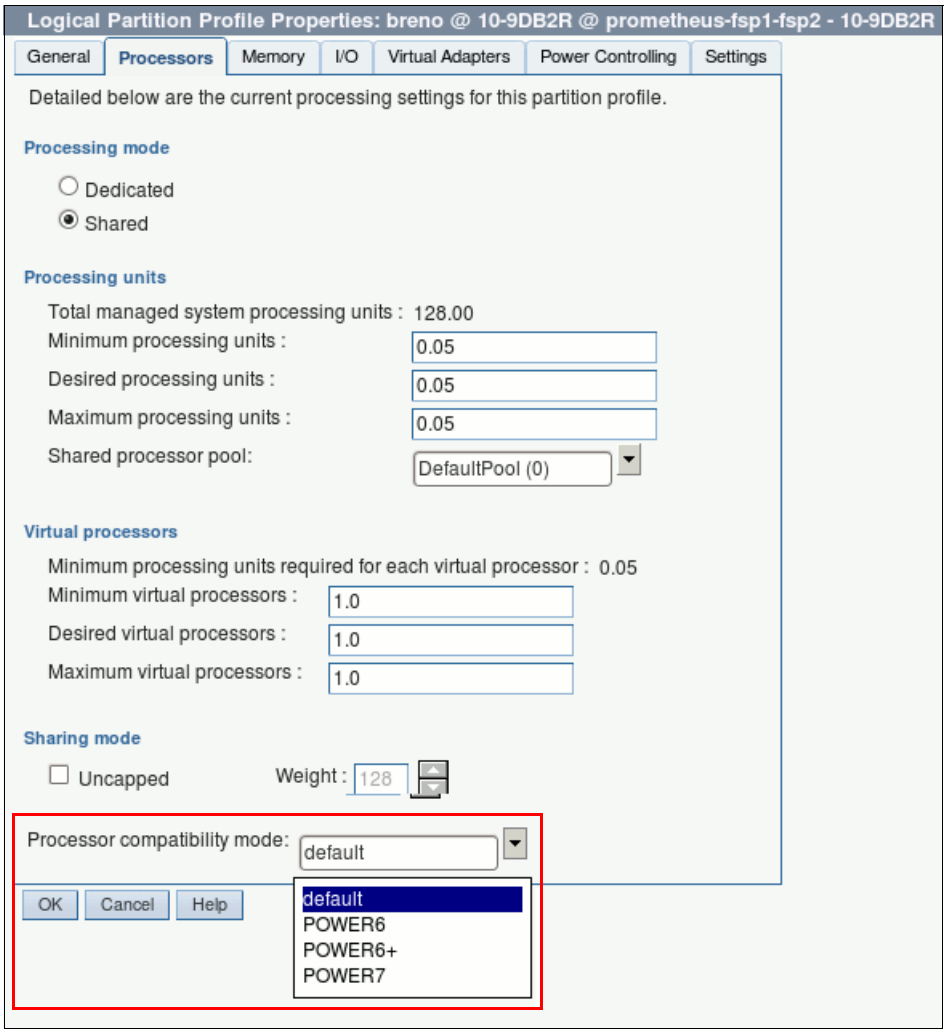


Figure 3-2 Configuring partition profile compatibility mode from the managed console

Table 3-2 lists the differences between these modes.

Table 3-2 Differences between POWER6, POWER6+, and POWER7 compatibility mode

POWER6 and POWER6+ mode	POWER7 mode	Customer value
2-thread SMT	4-thread SMT	Throughput performance, processor core utilization
Vector Multimedia Extension/ AltiVec (VMX)	Vector scalar extension (VSX)	High-performance computing
Affinity off by default	3-tier memory, micropartition affinity, dynamic platform optimizer	Improved system performance for system images spanning sockets and nodes
► Barrier synchronization ► Fixed 128-byte array, kernel extension access	► Enhanced barrier Synchronization ► Variable sized array, user shared memory access	High-performance computing parallel programming synchronization facility
64-core and 128-thread scaling	► 32-core and 128-thread scaling ► 64-core and 256-thread scaling ► 128-core and 512-thread scaling ► 256-core and 1024-thread scaling	Performance and scalability for large scale-up single system image workloads (such as OLTP, ERP scale-up, and WPAR consolidation)
EnergyScale CPU Idle	EnergyScale CPU Idle and Folding with NAP and SLEEP	Improved energy efficiency

3.3 Active Memory Expansion

Active Memory Expansion enablement is an optional feature of POWER7 and POWER7+ processor-based servers that must be specified by using FC 4795 when creating the configuration in the e-Config tool.

This feature enables memory expansion on the system. By using compression and decompression of memory content can effectively expand the maximum memory capacity, providing additional server workload capacity and performance.

Active Memory Expansion is a POWER technology that allows the effective maximum memory capacity to be much larger than the true physical memory maximum. Compression and decompression of memory content can allow memory expansion up to 125% for AIX partitions, which in turn enables a partition to perform significantly more work or support more users with the same physical amount of memory. Similarly, it can allow a server to run more partitions and do more work for the same physical amount of memory.

Active Memory Expansion is available for partitions running AIX 6.1, Technology Level 4 with SP2, or later.

Active Memory Expansion uses the CPU resource of a partition to compress and decompress the memory contents of this same partition.The trade-off of memory capacity for processor cycles can be an excellent choice, but the degree of expansion varies based on how

compressible the memory content is, and it also depends on having adequate spare CPU capacity available for this compression and decompression.

The POWER7+ processor includes Active Memory Expansion on the processor chip to provide dramatic improvement in performance and greater processor efficiency. To take advantage of the hardware compression offload, AIX 6.1 Technology Level 8 is required.

The Active Memory Expansion feature is not supported with the IBM i and Linux operating systems.

Tests in IBM laboratories, using sample work loads, showed excellent results for many workloads in terms of memory expansion per additional CPU utilized. Other test workloads had more modest results. The ideal scenario is when there are many cold pages, that is, infrequently referenced pages. However, if many memory pages are referenced frequently, the Active Memory Expansion might not be a good choice.

Tip: If the workload is Java-based, the garbage collector must be tuned, so that it does not access the memory pages so often, turning cold pages to hot.

Clients have much control over Active Memory Expansion usage. Each individual AIX partition can turn on or turn off Active Memory Expansion. Control parameters set the amount of expansion you want in each partition to help control the amount of CPU that is used by the Active Memory Expansion function. An initial program load (IPL) is required for the specific partition that is turning memory expansion on or off. After turned on, monitoring capabilities are available in standard AIX performance tools, such as **lparstat**, **vmstat**, **topas**, and **svmon**. For specific POWER7+ hardware compression, the **amepat** tool is used to configure the offload details.

Figure 3-3 represents the percentage of CPU that is used to compress memory for two partitions with separate profiles. Curve 1 corresponds to a partition that has spare processing power capacity. Curve 2 corresponds to a partition that is constrained in processing power.

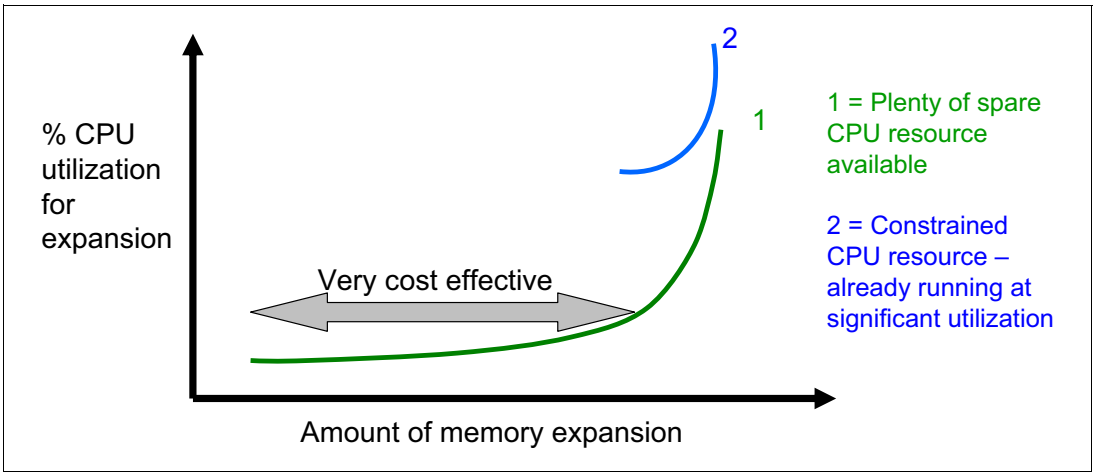


Figure 3-3 CPU usage versus memory expansion effectiveness

Both cases show that there is a “knee-of-curve” relationship for the CPU resource required for memory expansion:

- Busy processor cores do not have resources to spare for expansion.
- The more memory expansion is done, the more CPU resource is required.

The knee varies depending on how compressible the memory contents are. This example demonstrates the need for a case-by-case study of whether memory expansion can provide a positive return on investment.

To help you do this study, a planning tool is included with AIX 6.1 Technology Level 4 SP2, allowing you to sample actual workloads and estimate how expandable the partition's memory is and how much CPU resource is needed. Any model Power System can run the planning tool.

Figure 3-4 shows an example of the output that is returned by this planning tool. The tool outputs various real memory and CPU resource combinations to achieve the desired effective memory. It also recommends one particular combination. In this example, the tool recommends that you allocate 13% of processing power (2.13 physical processors in this setup) to benefit from 119% extra memory capacity.

```
Active Memory Expansion Modeled Statistics:
-------------------------------------------
Modeled Expanded Memory Size : 52.00 GB
Achievable Compression ratio : 4.51

Expansion    Modeled True      Modeled              CPU Usage
Factor       Memory Size       Memory Gain          Estimate
---------    -------------     ------------------   -----------
    1.40         37.25 GB        14.75 GB [ 40%]   0.00 [  0%]
    1.80         29.00 GB        23.00 GB [ 79%]   0.87 [  5%]
    2.19         23.75 GB        28.25 GB [119%]   2.13 [ 13%]
    2.57         20.25 GB        31.75 GB [157%]   2.96 [ 18%]
    2.98         17.50 GB        34.50 GB [197%]   3.61 [ 23%]
    3.36         15.50 GB        36.50 GB [235%]   4.09 [ 26%]

Active Memory Expansion Recommendation:
---------------------------------------
The recommended AME configuration for this workload is to configure the LPAR
with a memory size of 23.75 GB and to configure a memory expansion factor
of 2.19.  This will result in a memory gain of 119%. With this
configuration, the estimated CPU usage due to AME is approximately 2.13
physical processors, and the estimated overall peak CPU resource required for
the LPAR is 11.65 physical processors.
```

Figure 3-4 Output from Active Memory Expansion planning tool

After you select the value of the memory expansion factor that you want to achieve, you can use this value to configure the partition from the managed console (Figure 3-5).

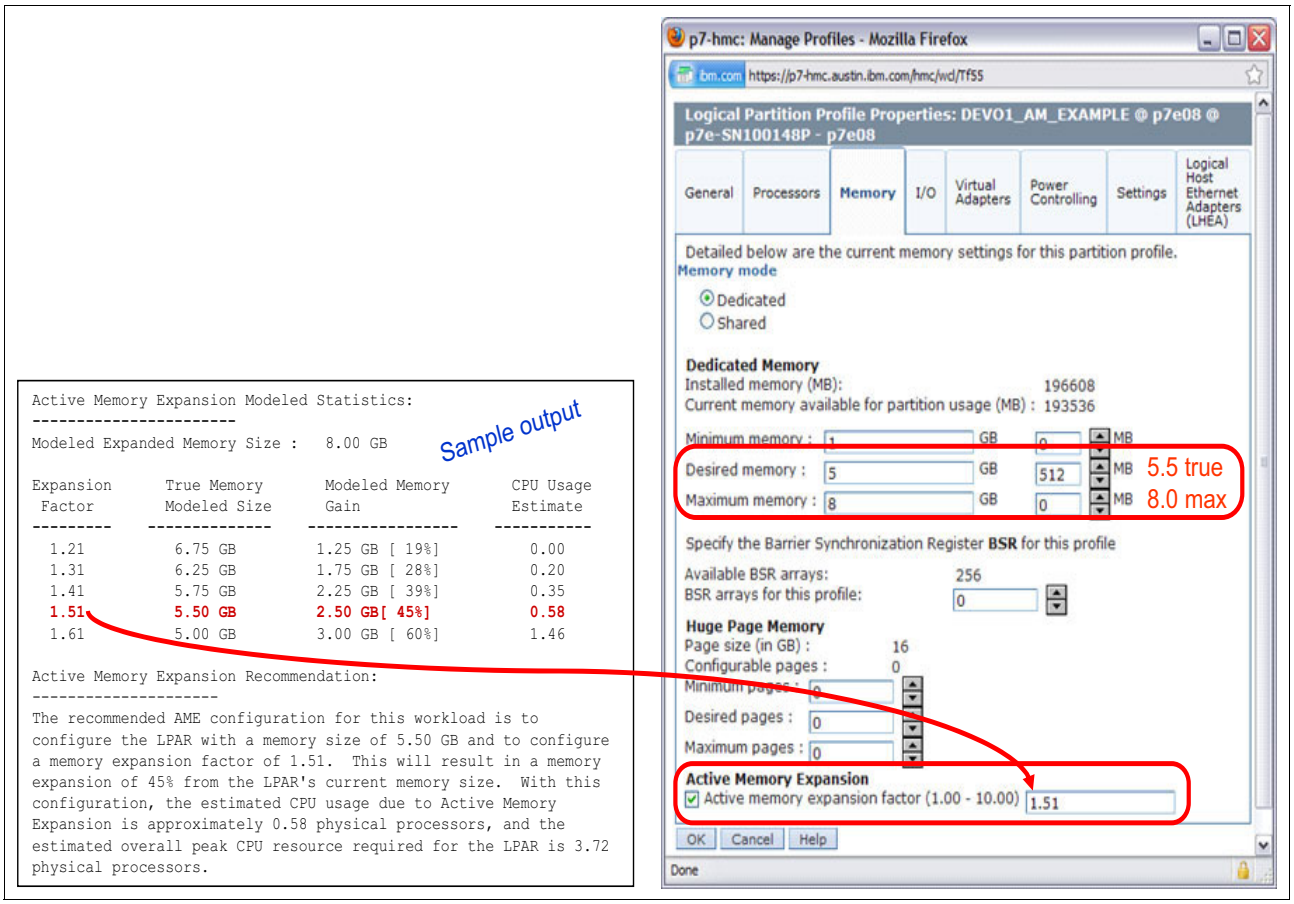


Figure 3-5 Using the planning tool result to configure the partition

On the HMC menu that describes the partition, select the **Active Memory Expansion** check box and enter the true and maximum memory, and the memory expansion factor. To turn off expansion, clear the check box. In both cases, reboot the partition to activate the change.

In addition, a one-time, 60-day trial of Active Memory Expansion is available to provide more exact memory expansion and CPU measurements. The trial can be requested by using the Power Systems Capacity on Demand web page:

http://www.ibm.com/systems/power/hardware/cod/

Active Memory Expansion can be ordered with the initial order of the server or as a miscellaneous equipment specification (MES) order. A software key is provided when the enablement feature is ordered that is applied to the server. Rebooting is not required to enable the physical server. The key is specific to an individual server and is permanent. It cannot be moved to a separate server. This feature is ordered per server, independent of the number of partitions using memory expansion.

From the HMC, you can view whether the Active Memory Expansion feature was activated (Figure 3-6).

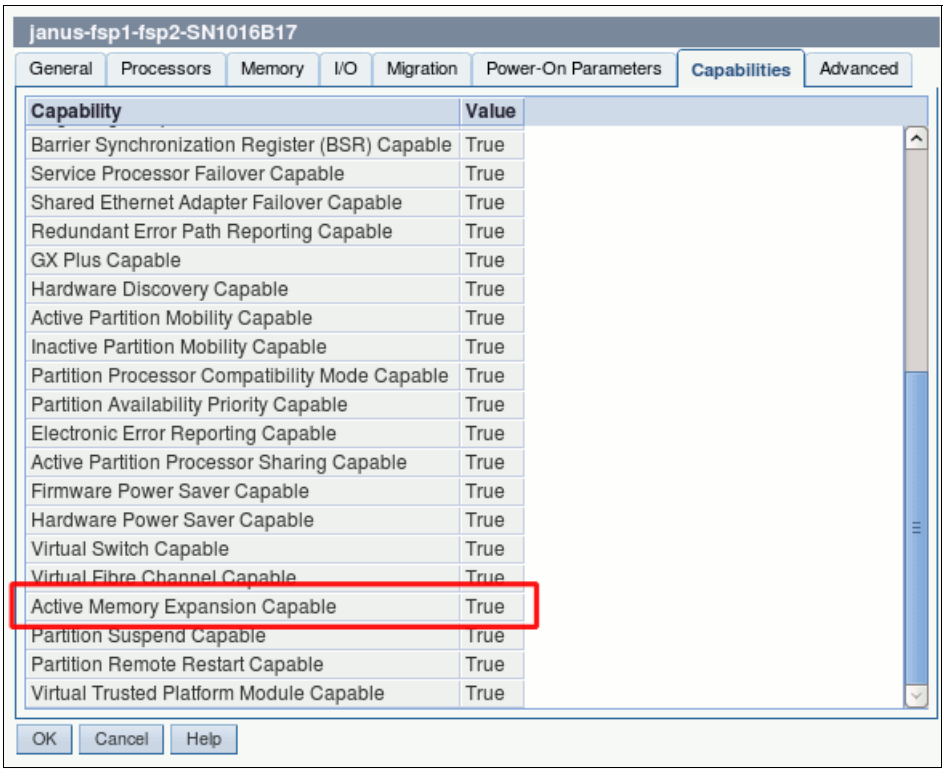


Figure 3-6 Server capabilities listed from the HMC

Moving an LPAR: If you want to move an LPAR that uses Active Memory Expansion to a system that uses Live Partition Mobility, the target system must support Active Memory Expansion (the target system must have Active Memory Expansion activated with the software key). If the target system does not have Active Memory Expansion activated, the mobility operation fails during the premobility check phase, and an appropriate error message is displayed.

For details about Active Memory Expansion, download the document *Active Memory Expansion: Overview and Usage Guide*:

http://public.dhe.ibm.com/common/ssi/ecm/en/pow03037usen/POW03037USEN.PDF

3.4 PowerVM

The PowerVM platform is the family of technologies, capabilities, and offerings that delivers industry-leading virtualization on the IBM Power Systems. It is the umbrella branding term for Power Systems virtualization (Logical Partitioning, IBM Micro-Partitioning®, POWER Hypervisor, Virtual I/O Server, Live Partition Mobility, Workload Partitions, and more). As with Advanced Power Virtualization in the past, PowerVM is a combination of hardware enablement and value-added software. The licensed features of each of the three separate editions of PowerVM are described in 3.4.1, “PowerVM editions” on page 114.

3.4.1 PowerVM editions

The three editions of PowerVM are suited for various purposes:

- PowerVM Express Edition

  This edition is designed for customers who want an introduction to more advanced virtualization features at a highly affordable price, generally in single-server projects.

- PowerVM Standard Edition

  This edition provides advanced virtualization functions and is intended for production deployments and server consolidation.

- PowerVM Enterprise Edition

  This edition is suitable for large server deployments such as multi-server deployments and cloud infrastructures. It includes unique features such as Active Memory Sharing and Live Partition Mobility.

Table 3-3 lists the editions of PowerVM that are available on Power 710 and Power 730.

Table 3-3 Availability of PowerVM per POWER7+ processor technology-based server model

Servers	Express	Standard	Enterprise
IBM Power 710	FC 5225	FC 5227	FC 5228
IBM Power 730	FC 5225	FC 5227	FC 5228

For more information about the features included on each version of PowerVM, see *IBM PowerVM Virtualization Introduction and Configuration*, SG24-7940-04.

3.4.2 Logical partitions

Logical partitions (LPARs) and virtualization increase use of system resources and add a new level of configuration possibilities.

Logical partitioning

Logical partitioning was introduced with the POWER4 processor-based product line and the AIX Version 5.1, Red Hat Enterprise Linux 3.0 and SUSE Linux Enterprise Server 9.0 operating systems. This technology offered the capability to divide a pSeries system into separate logical systems, allowing each LPAR to run an operating environment on dedicated attached devices, such as processors, memory, and I/O components.

Later, dynamic logical partitioning increased the flexibility, allowing selected system resources, such as processors, memory, and I/O components, to be added and deleted from logical partitions while they are executing. AIX Version 5.2, with all the necessary enhancements to enable dynamic LPAR, was introduced in 2002. At the same time, Red Hat Enterprise Linux 5 and SUSE Linux Enterprise 9.0 were also able do support dynamic logical partitioning. 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.

Micro-Partitioning

The IBM 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 micropartition. Micropartitions run over a set of processors called a *shared processor pool*,

and 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. The meaning of the term *physical processor* in this section is a *processor core*. For example, a 2-core server has two physical processors.

When defining a shared processor partition, several options must be defined:

- The minimum, desired, and maximum processing units

  Processing units are defined as processing power, or the fraction of time that the partition is dispatched on physical processors. Processing units define the capacity entitlement of the partition.

- The shared processor pool

  Select one from the list with the names of each configured shared processor pool. This list also displays, in parentheses, the pool ID of each configured shared processor pool. If the name of the desired shared processor pool is not available here, you must first configure the shared processor pool by using the shared processor pool Management window. Shared processor partitions use the default shared processor pool, called DefaultPool by default. See 3.4.3, “Multiple shared processor pools” on page 117, for details about multiple shared processor pools.

- Whether the partition will be able to access extra processing power to “fill up” its virtual processors above its capacity entitlement (selecting either to cap or uncap your partition).

  If spare processing power is available in the shared processor pool or other partitions are not using their entitlement, an uncapped partition can use additional processing units if its entitlement is not enough to satisfy its application processing demand.

- The weight (preference) in the case of an uncapped partition.

- The minimum, desired, and maximum number of virtual processors.

The POWER Hypervisor calculates partition processing power based on minimum, desired, and maximum values, processing mode, and is also based on requirements of other active partitions. 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.

On the POWER7+ processors, a partition can be defined with a processor capacity as small as 0.05 processing units. This number represents 0.05 of a physical processor. Each physical processor can be shared by up to 20 shared processor partitions, and the partition’s entitlement can be incremented fractionally by as little as 0.01 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.

IBM Power 710 supports up to eight cores, and these maximum numbers:

- 8 dedicated partitions
- 160 micropartitions (maximum 20 micropartitions per physical active core)

The Power 730 supports up to 16 cores in a single system, and these maximum numbers:

- 16 dedicated partitions
- 320 micropartitions (maximum 20 micropartitions per physical active core)

An important point is that the maximum amounts are supported by the hardware, but the practical limits depend on application workload demands.

Consider the following additional information about virtual processors:

- A virtual processor can be running (dispatched) either on a physical processor or as standby waiting for a physical processor to became available.
- Virtual processors do not introduce any additional abstraction level. They 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.

Processing mode

When you create a logical partition, you can assign entire processors for dedicated use, or you can assign partial processing units from a shared processor pool. This setting defines the processing mode of the logical partition. Figure 3-7 shows a diagram of the concepts described in this section.

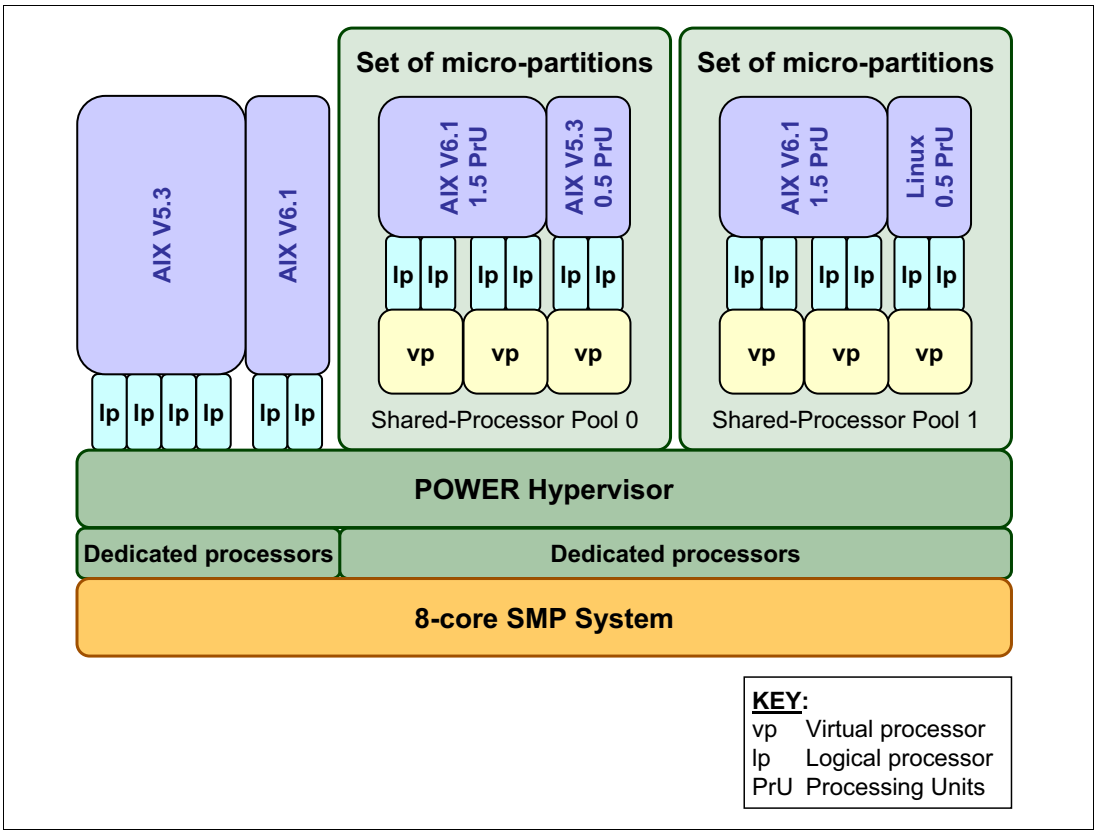


Figure 3-7 Logical partitioning concepts

Dedicated mode

In dedicated mode, physical processors are assigned as a whole to partitions. The simultaneous multithreading feature in the POWER7+ processor core allows the core to execute instructions from two or four independent software threads simultaneously. To support this feature consider the concept of *logical processors*. The operating system (AIX, IBM i, or Linux) sees one physical processor as two or four logical processors if the simultaneous multithreading feature is on. It can be turned off and on dynamically while the operating system is executing (for AIX, use the **smtctl** command; for Linux, use the **ppc64_cpu --smt** command). If simultaneous multithreading is off, each physical processor is presented as one logical processor, and thus only one thread.

Shared dedicated mode

On POWER7+ processor technology-based servers, you can configure dedicated partitions to become processor donors for idle processors that 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.

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 that are assigned to the partition). From the POWER Hypervisor perspective, virtual processors represent dispatching objects. The POWER Hypervisor dispatches virtual processors to physical processors according to the 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 receive total CPU time equal to their processing unit's entitlement. The logical processors are defined on top of virtual processors. So, even with a virtual processor, the concept of a logical processor exists and the number of logical processors depends whether the simultaneous multithreading is turned on or off.

3.4.3 Multiple shared processor pools

Multiple shared processor pools (MSPPs) is a capability that is supported on POWER6, POWER6+, POWER7, and POWER7+ processor-based servers. This capability allows a system administrator to create a set of micropartitions with the purpose of controlling the processor capacity that can be consumed from the physical shared processor pool.

Implementing MSPPs depends on a set of underlying techniques and technologies. Figure 3-8 is an overview of the architecture of multiple shared processor pools.

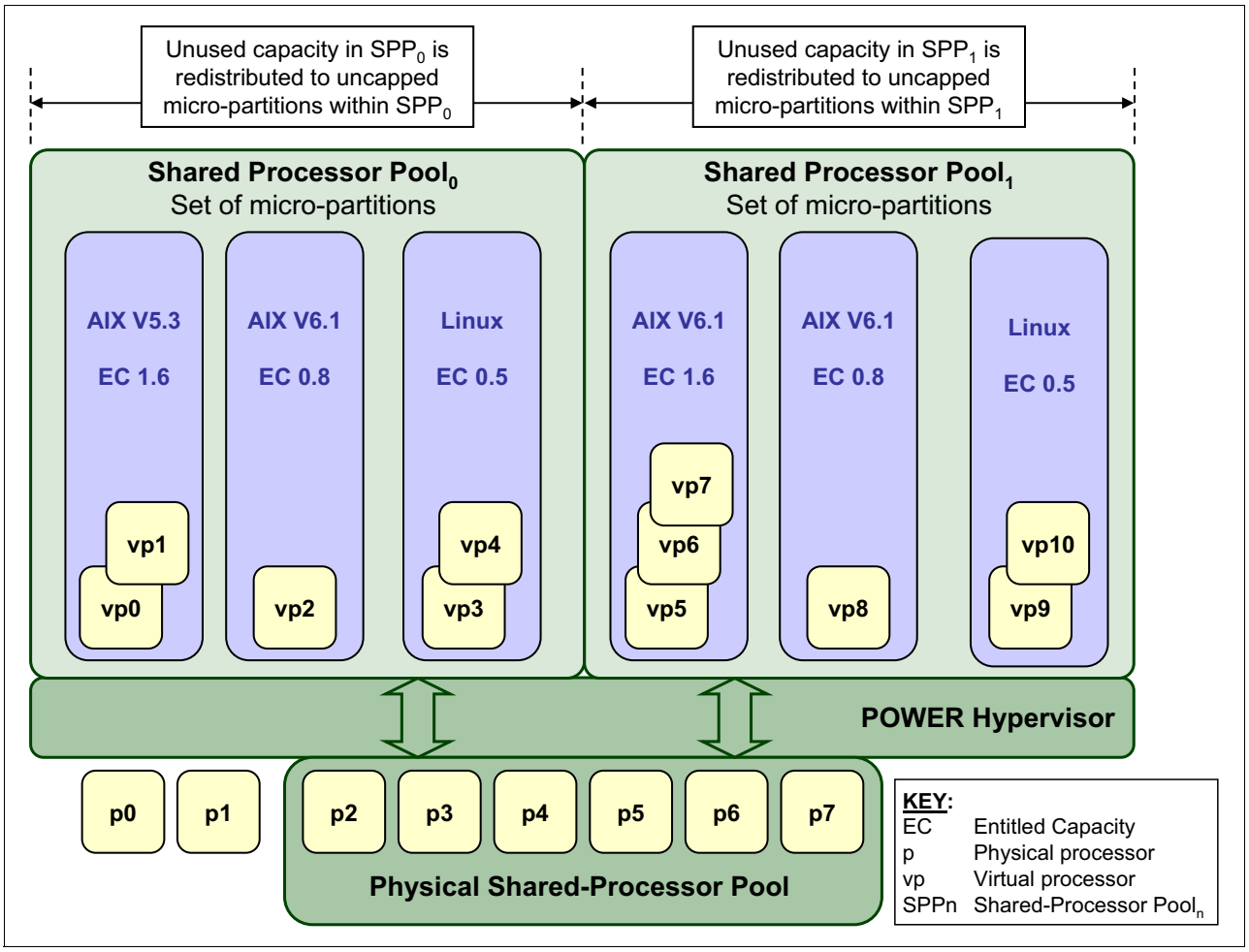


Figure 3-8 Overview of the architecture of multiple shared processor pools

Micropartitions are created and then identified as members of either the default shared processor pool$_0$ or a user-defined shared processor pool$_n$. The virtual processors that exist within the set of micropartitions are monitored by the POWER Hypervisor, and processor capacity is managed according to user-defined attributes.

If the Power Systems server is under heavy load, each micropartition within a shared processor pool is guaranteed its processor entitlement plus any capacity that it might be allocated from the reserved pool capacity if the micropartition is uncapped.

If certain micropartitions in a shared processor pool do not use their capacity entitlement, the unused capacity is ceded and other uncapped micropartitions within the same shared processor pool are allocated the additional capacity according to their uncapped weighting. In this way, the entitled pool capacity of a shared processor pool is distributed to the set of micropartitions within that shared processor pool.

All Power Systems servers that support the multiple shared processor pools capability have a minimum of one (the default) shared processor pool and up to a maximum of 64 shared processor pools.

Default shared processor pool (SPP_0)

On any Power Systems server supporting multiple shared processor pools, a default shared processor pool is always automatically defined. The default shared processor pool has a pool identifier of zero (SPP-ID = 0) and can also be referred to as SPP_0. The default shared processor pool has the same attributes as a user-defined shared processor pool except that these attributes are not directly under the control of the system administrator. They have fixed values (Table 3-4).

Table 3-4 Attribute values for the default shared processor pool (SPP_0)

SPP_0 attribute	**Value**
Shared processor pool ID	0
Maximum pool capacity	Value is equal to the capacity in the physical shared processor pool.
Reserved pool capacity	0
Entitled pool capacity	Sum (total) of the entitled capacities of the micropartitions in the default shared processor pool.

Creating multiple shared processor pools

The default shared processor pool (SPP_0) is automatically activated by the system and is always present.

All other shared processor pools exist, but by default are inactive. By changing the maximum pool capacity of a shared processor pool to a value greater than zero, it becomes active and can accept micropartitions (either transferred from SPP_0 or newly created).

Levels of processor capacity resolution

The following two levels of processor capacity resolution are implemented by the POWER Hypervisor and multiple shared processor pools:

- $Level_0$

  This first level is the resolution of capacity within the same shared processor pool. Unused processor cycles from within a shared processor pool are harvested and then redistributed to any eligible micropartition within the same shared processor pool.

- $Level_1$

  This second level is after all first level capacity is resolved. When all $Level_0$ capacity has been resolved within the multiple shared processor pools, the POWER Hypervisor harvests unused processor cycles and redistributes them to eligible micropartitions regardless of the multiple shared processor pools structure.

Figure 3-9 shows the levels of unused capacity redistribution that are implemented by the POWER Hypervisor.

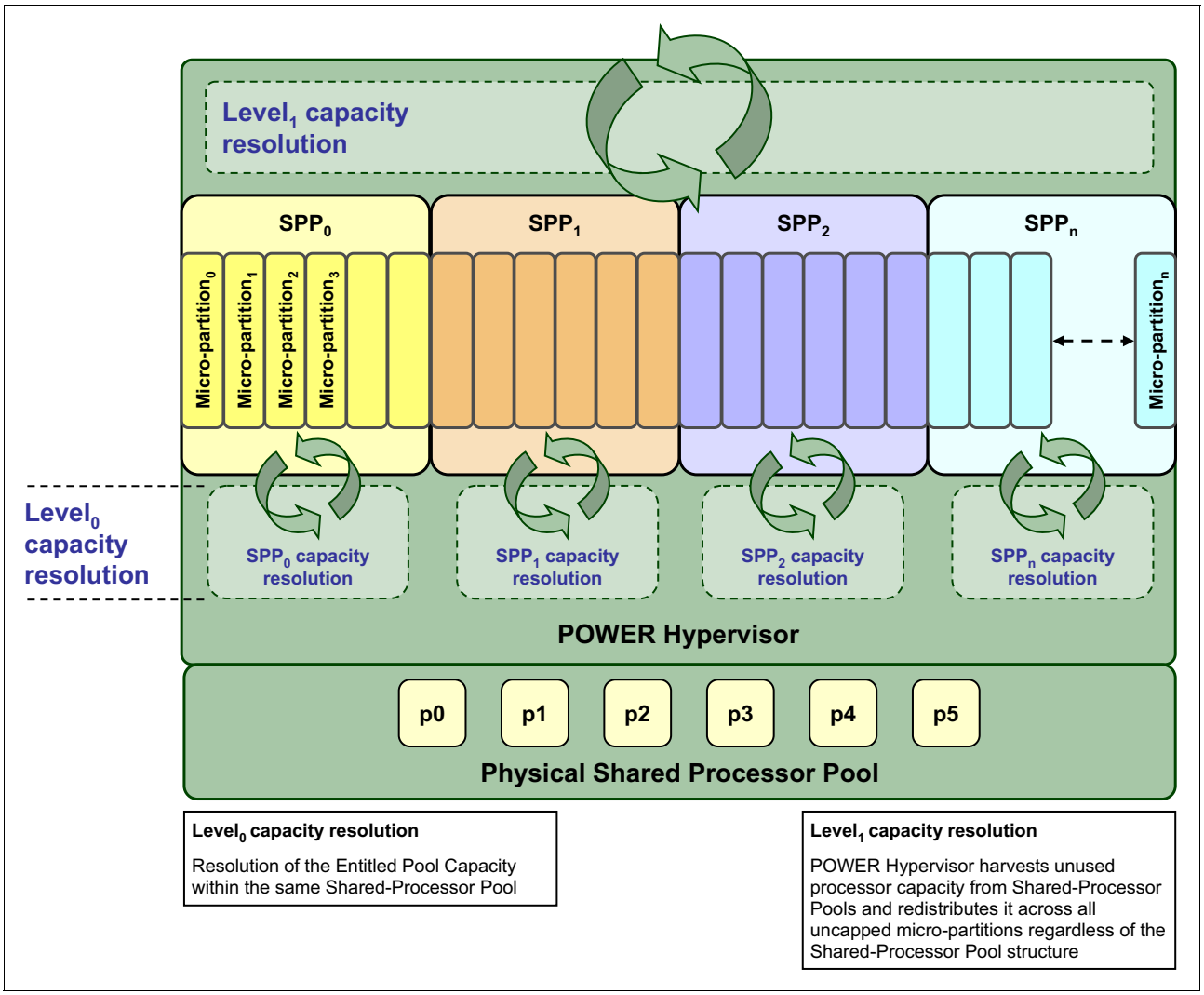


Figure 3-9 The levels of unused capacity redistribution

Capacity allocation above the entitled pool capacity (Level$_1$)

The POWER Hypervisor initially manages the entitled pool capacity at the shared processor pool level. This level is where unused processor capacity within a shared processor pool is harvested and then redistributed to uncapped micropartitions within the same shared processor pool. This level of processor capacity management is sometimes referred to as Level$_0$ capacity resolution.

At a higher level, the POWER Hypervisor harvests unused processor capacity from the multiple shared processor pools that do not consume all of their entitled pool capacity. If a particular shared processor pool is heavily loaded and several of the uncapped micropartitions within it require additional processor capacity (above the entitled pool capacity), then the POWER Hypervisor redistributes some of the extra capacity to the uncapped micropartitions. This level of processor capacity management is sometimes referred to as Level$_1$ capacity resolution.

To redistribute unused processor capacity to uncapped micropartitions in multiple shared processor pools above the entitled pool capacity, the POWER Hypervisor uses a higher level of redistribution, $Level_1$.

$Level_1$ capacity resolution: When allocating additional processor capacity in excess of the entitled pool capacity of the shared processor pool, the POWER Hypervisor takes the uncapped weights of *all micropartitions in the system* into account, *regardless of the multiple shared processor pool structure*.

Where there is unused processor capacity in under-utilized shared processor pools, the micropartitions within the shared processor pools cede the capacity to the POWER Hypervisor.

In busy shared processor pools, where the micropartitions used all of the entitled pool capacity, the POWER Hypervisor allocates additional cycles to micropartitions, in which *all* of the following statements are true:

- The maximum pool capacity of the shared processor pool that is hosting the micropartition is not met.
- The micropartition is uncapped.
- The micropartition has enough virtual-processors to take advantage of the additional capacity.

Under these circumstances, the POWER Hypervisor allocates additional processor capacity to micropartitions on the basis of their uncapped weights, independent of the shared processor pool that hosts the micropartitions. This behavior can be referred to as $Level_1$ capacity resolution. Consequently, when allocating additional processor capacity in excess of the entitled pool capacity of the shared processor pools, the POWER Hypervisor takes the uncapped weights of all micropartitions in the system into account, regardless of the multiple shared processor pool structure.

Dynamic adjustment of maximum pool capacity

The maximum pool capacity of a shared processor pool, other than the default shared processor $pool_0$, can be adjusted dynamically from the managed console, using either the graphical interface or the command-line interface (CLI).

Dynamic adjustment of reserved pool capacity

The reserved pool capacity of a shared processor pool, other than the default shared processor $pool_0$, can be adjusted dynamically from the managed console, by using either the graphical interface or the CLI.

Dynamic movement between shared processor pools

A micropartition can be moved dynamically from one shared processor pool to another by using the managed console with either the graphical interface or the CLI. Because the entitled pool capacity is partly made up of the sum of the entitled capacities of the micropartitions, removing a micropartition from a shared processor pool reduces the entitled pool capacity for that shared processor pool. Similarly, the entitled pool capacity of the shared processor pool that the micropartition joins will increase.

Deleting a shared processor pool

Shared processor pools cannot be deleted from the system. However, they are deactivated by setting the maximum pool capacity and the reserved pool capacity to zero. The shared processor pool will still exist but will not be active. Use the managed console interface to deactivate a shared processor pool. A shared processor pool cannot be deactivated unless all micropartitions hosted by the shared processor pool have been removed.

Live Partition Mobility and multiple shared processor pools

A micropartition can leave a shared processor pool because of PowerVM Live Partition Mobility. Similarly, a micropartition can join a shared processor pool in the same way. When performing PowerVM Live Partition Mobility, you are given the opportunity to designate a destination shared processor pool on the target server to receive and host the migrating micropartition.

Because several simultaneous micropartition migrations are supported by PowerVM Live Partition Mobility, migrating the entire shared processor pool from one server to another is conceivable.

3.4.4 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 3-10 shows an overview of a Virtual I/O Server configuration.

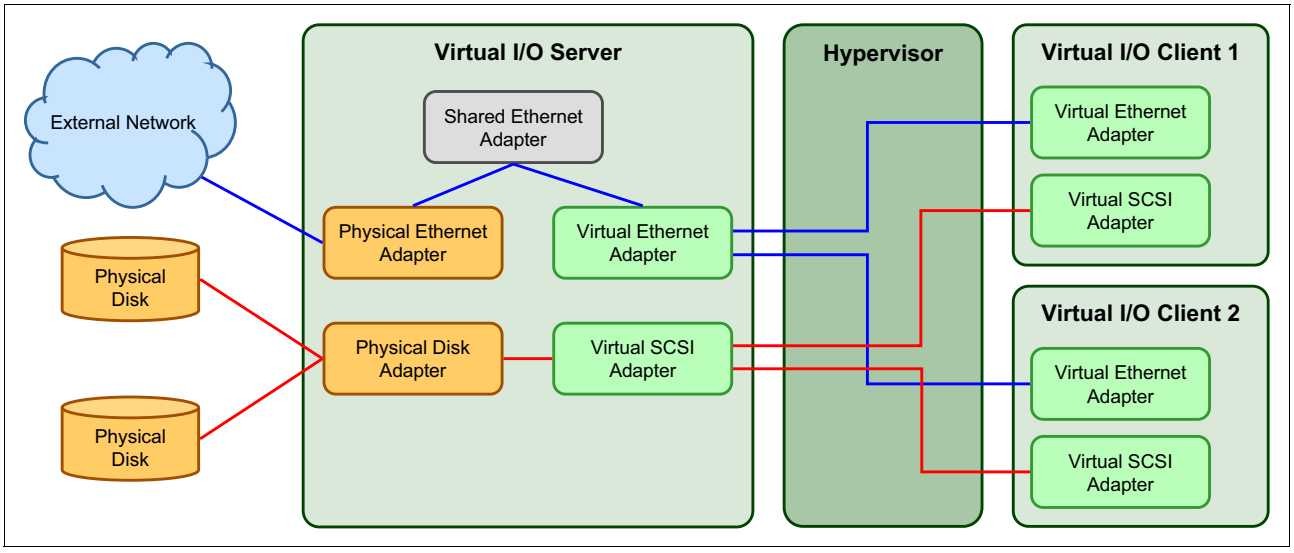


Figure 3-10 Architectural view of the Virtual I/O Server

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 who order any PowerVM edition. This dedicated software is only for the Virtual I/O Server (and IVM in case it is used) and is supported only in special Virtual I/O Server partitions. Three major virtual devices are supported by the Virtual I/O Server:

- Shared Ethernet Adapter
- Virtual SCSI
- Virtual Fibre Channel adapter

The Virtual Fibre Channel adapter is used with the NPIV feature, described in 3.4.10, “Operating system support for PowerVM” on page 134.

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. With an SEA, you can connect internal and external VLANs by using a physical adapter. The Shared Ethernet Adapter service can be hosted only 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 also, by using the **`brctl`** command.

Figure 3-11 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.

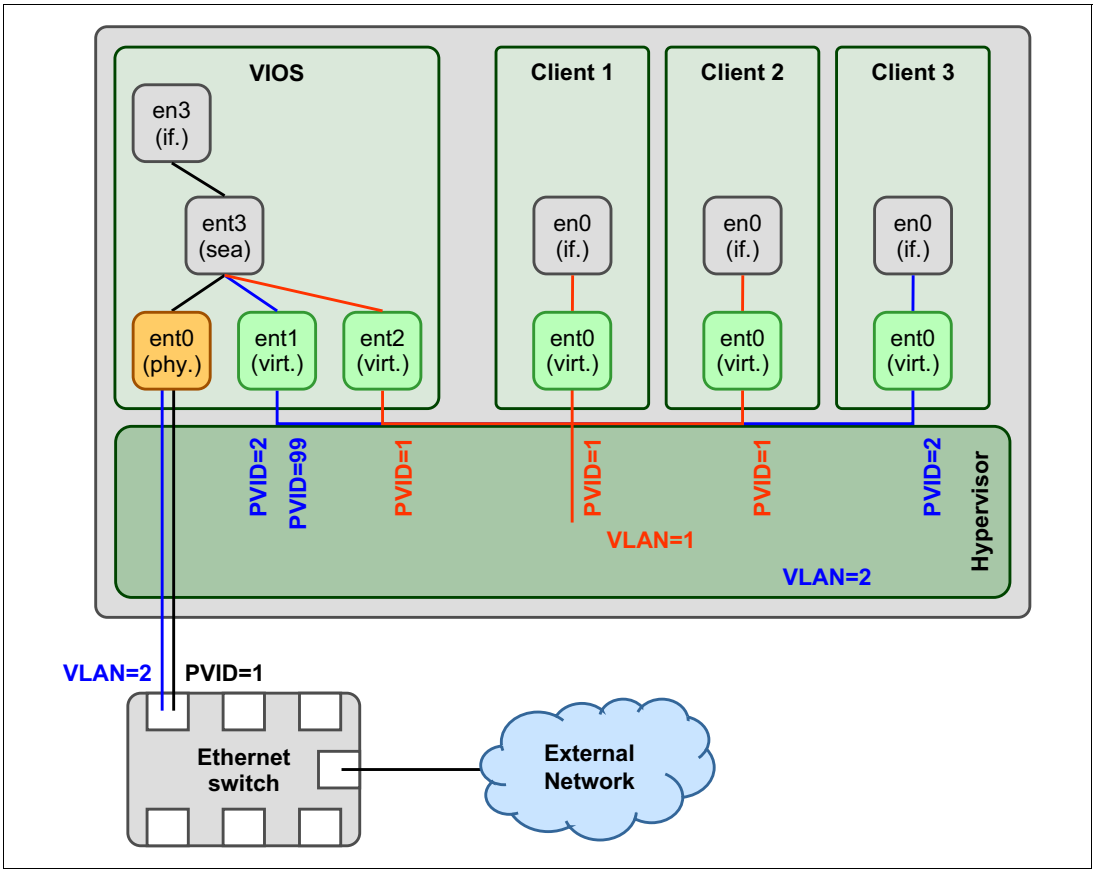


Figure 3-11 Architectural view of a Shared Ethernet Adapter

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, a possibility is for a single physical Ethernet to be shared between 320 internal VLAN networks. 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 on an SEA.

IP address: A Shared Ethernet Adapter does not require a configured IP address to be able to perform the Ethernet bridging functionality. Configuring IP on the Virtual I/O Server is convenient 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 task 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 task leaves the SEA without the IP address, allowing for maintenance on the SEA without losing IP connectivity in case SEA failover is configured.

Virtual SCSI

Virtual SCSI is used to see 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 a server or, in SCSI terms, a 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 a managed console or through the Integrated Virtualization Manager on smaller systems. The virtual SCSI server (target) adapter is responsible for executing any SCSI commands that it receives. It is owned by the Virtual I/O Server partition. The virtual SCSI client adapter allows a client partition to access physical SCSI 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 that are presented to the Virtual/O Server can be exported and assigned to a client partition in various 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 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 separate partitions. Therefore, virtual SCSI enables sharing of adapters and disk devices.

Figure 3-12 shows an example where one physical disk is divided into two logical volumes by the Virtual I/O Server. Each client partition 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*.

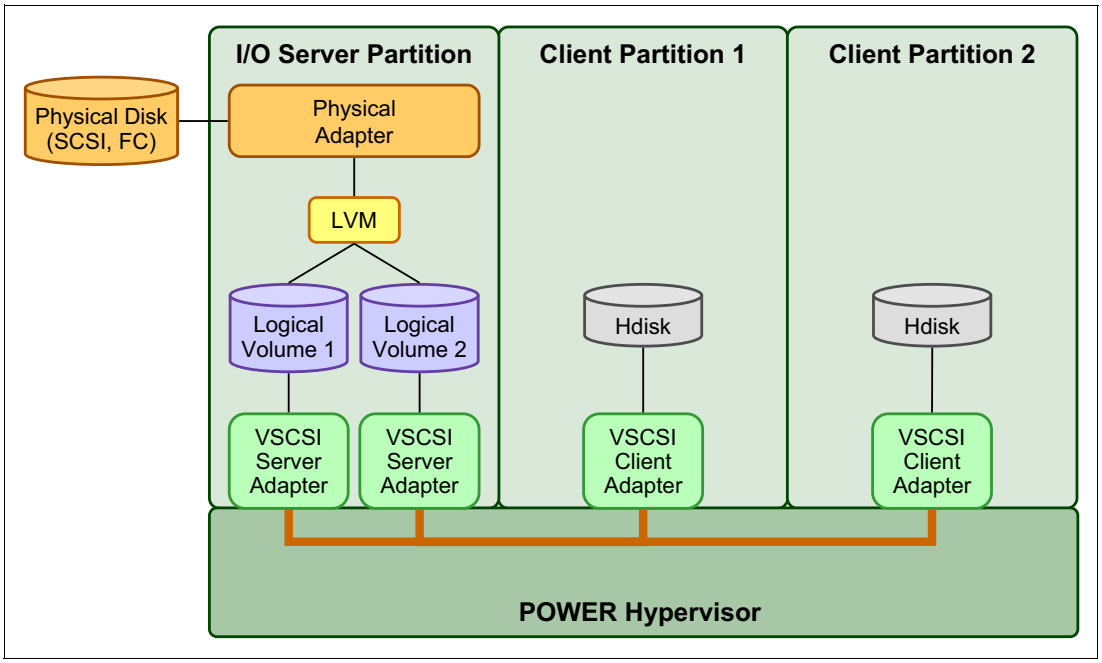


Figure 3-12 Architectural view of virtual SCSI

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 specific storage devices that are supported for Virtual I/O Server, see the following web page:

http://www14.software.ibm.com/webapp/set2/sas/f/vios/documentation/datasheet.html

N_Port ID Virtualization

N_Port ID Virtualization (NPIV) is a technology that allows multiple logical partitions to access independent physical storage through the same physical Fibre Channel adapter. This adapter is attached to a Virtual I/O Server partition that acts only as a pass-through, managing the data transfer through the POWER Hypervisor.

Each partition that uses NPIV is identified by a pair of unique worldwide port names, enabling you to connect each partition to independent physical storage on a SAN. Unlike virtual SCSI, only the client partitions see the disk.

For more information and requirements for NPIV, see the following resources:

- *PowerVM Migration from Physical to Virtual Storage*, SG24-7825
- *IBM PowerVM Virtualization Managing and Monitoring*, SG24-7590

Virtual I/O Server functions

The Virtual I/O Server has many features, including monitoring solutions:

- Support for Live Partition Mobility starting on POWER6 processor-based systems with the PowerVM Enterprise Edition. For more information about Live Partition Mobility, see 3.4.5, “PowerVM Live Partition Mobility” on page 127.
- Support for virtual SCSI devices backed by a file, which are then accessed as standard SCSI-compliant LUNs.
- Support for virtual Fibre Channel devices that are used with the NPIV feature.
- Virtual I/O Server Expansion Pack with additional security functions such as Kerberos (Network Authentication Service for users and client and server applications), Simple Network Management Protocol (SNMP) v3, and Lightweight Directory Access Protocol (LDAP) client functionality.
- System Planning Tool (SPT) and Workload Estimator, which are designed to ease the deployment of a virtualized infrastructure. For more information about the System Planning Tool, see 3.5, “System Planning Tool” on page 137.
- IBM Systems Director agent and several preinstalled IBM Tivoli® agents, such as the following examples:
  - Tivoli Identity Manager, to allow easy integration into an existing Tivoli Systems Management infrastructure
  - Tivoli Application Dependency Discovery Manager (ADDM), which creates and automatically maintains application infrastructure maps including dependencies, change-histories, and deep configuration values
- vSCSI enterprise reliability, availability, serviceability (eRAS).
- Additional CLI statistics in **svmon**, **vmstat**, **fcstat**, and **topas**.
- Monitoring solutions to help manage and monitor the Virtual I/O Server and shared resources. Commands and views provide additional metrics for memory, paging, processes, Fibre Channel HBA statistics, and virtualization.

For more information about the Virtual I/O Server and its implementation, see *PowerVM Virtualization on IBM System p: Introduction and Configuration Fourth Edition*, SG24-7940.

3.4.5 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 or 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.

Live Partition Mobility provides systems management flexibility and improves system availability:

- 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 approach allows your users to continue their work without disruption.
- Avoid unplanned downtime. With 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.
- Take advantage of server optimization:
  - Consolidation: You can consolidate workloads that run on several small, under-used servers onto a single large server.
  - Deconsolidation: You can move 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.

Hardware and operating system requirements for Live Partition Mobility

PowerVM Live Partition Mobility requires systems with POWER6 or newer processors to run PowerVM Enterprise Edition and is supported for partitions running the following levels of operating systems:

- AIX 5.3 TL7 or later
- IBM i 7.1 TR4 or later
- SUSE Linux Enterprise Server 10 Service Pack 4 or later
- Red Hat Enterprise Linux version 5 Update 1 or later

The Virtual I/O Server partition itself cannot be migrated.

Requirement for IBM i: Live Partition Mobility on IBM i is not supported on POWER6 or POWER6+-based servers.

Source and destination system requirements

The source partition must be one that has only virtual devices. If there are any physical devices in its allocation, they must be removed before the validation or migration is initiated. An N_Port ID Virtualization (NPIV) device is considered virtual and is compatible with partition migration.

The hypervisor must support the Live Partition Mobility functionality (also called migration process) that is available on POWER6, POWER6+, POWER7 and POWER7+

processor-based hypervisors. Firmware must be at firmware level eFW3.2 or later. All POWER7+ processor-based hypervisors support Live Partition Mobility. Source and destination systems can have separate firmware levels, but they must be compatible with each other.

A possibility is to migrate partitions back and forth between POWER6, POWER6+, POWER7 and POWER7+ processor-based servers. Partition Mobility uses the POWER6 or POWER6+ Compatibility Modes that are provided by POWER7 and POWER7+ processor-based servers. On the POWER7+ processor-based server, the migrated partition is then executing in POWER6 or POWER6+ Compatibility Mode.

Support of both processors: Because POWER7+ and POWER7 use the same Instruction Set Architecture (ISA), they are equivalent regarding partition mobility, that is POWER7 Compatibility Mode supports both POWER7 and POWER7+ processors.

If you want to move an active logical partition from a POWER6 processor-based server to a POWER7+ processor-based server so that the logical partition can take advantage of the additional capabilities available with the POWER7+ processor, use the following steps:

1. Set the partition-preferred processor compatibility mode to the default mode. When you activate the logical partition on the POWER6 or POWER6+ processor-based server, it runs in the POWER6 or POWER6+ mode.
2. Move the logical partition to the POWER7+ processor-based server. Both the current and preferred modes remain unchanged for the logical partition until you restart the logical partition.
3. Restart the logical partition on the POWER7+ processor-based server. The hypervisor evaluates the configuration. Because the preferred mode is set to default and the logical partition now runs on a POWER7+ processor-based server, the highest mode available is the POWER7+ mode. The hypervisor determines that the most fully featured mode that is supported by the operating environment that is installed in the logical partition is the POWER7 mode and changes the current mode of the logical partition to the POWER7 mode.

Now the current processor compatibility mode of the logical partition is the POWER7 mode, and the logical partition runs on the POWER7+ processor-based server.

Tip: The following web page offers presentations of the supported migrations:

`http://pic.dhe.ibm.com/infocenter/powersys/v3r1m5/index.jsp?topic=/p7hc3/iphc3pcmcombosact.htm`

The Virtual I/O Server on the source system provides the access to the client resources and must be identified as a mover service partition (MSP). The Virtual Asynchronous Services Interface (VASI) device allows the mover service partition to communicate with the hypervisor. It is created and managed automatically by the managed console and will be configured on both the source and destination Virtual I/O Servers, which are 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 must not be running on battery power, and shared storage (external hdisk with `reserve_policy=no_reserve`). In addition, all logical partitions must be on the same open network with RMC established to the managed console.

The managed console is used to configure, validate, and orchestrate. You use the managed console to configure the Virtual I/O Server as an MSP and to configure the VASI device. A managed console wizard validates your configuration and identifies issues that can cause

the migration to fail. During the migration, the managed console controls all phases of the process.

Improved Live Partition Mobility benefits

The possibility to move partitions between POWER6, POWER6+, POWER7, and POWER7+ processor-based servers greatly facilitates the deployment of POWER7+ processor-based servers, as follows:

- Installation of the new server can be done while the application is executing on a POWER6, POWER6+, or POWER7 server. After the POWER7+ processor-based server is ready, the application can be migrated to its new hosting server without application down time.
- When adding POWER7+ processor-based servers to a POWER6, POWER6+, and POWER7 environment, you have the additional flexibility to perform workload balancing across the entire set of POWER6, POWER6+, POWER7, and POWER7+ processor-based servers.
- When doing server maintenance, you have the additional flexibility to use POWER7 Servers for hosting applications, usually hosted on POWER7+ processor-based servers, allowing you to perform this maintenance with no interruption to application availability.

For more information about Live Partition Mobility and how to implement it, see *IBM PowerVM Live Partition Mobility*, SG24-7460.

3.4.6 Active Memory Sharing

Active Memory Sharing is an IBM PowerVM advanced memory virtualization technology that provides system memory virtualization capabilities to IBM Power Systems, allowing multiple partitions to share a common pool of physical memory.

Active Memory Sharing is available only with the Enterprise version of PowerVM.

The physical memory of an IBM Power System can be assigned to multiple partitions in either dedicated or shared mode. The system administrator has the capability to assign some physical memory to a partition and some physical memory to a pool that is shared by other partitions. A single partition can have either dedicated or shared memory:

- With a pure dedicated memory model, the system administrator's task is to optimize available memory distribution among partitions. When a partition suffers degradation because of memory constraints and other partitions have unused memory, the administrator can manually issue a dynamic memory reconfiguration.
- With a shared memory model, the system automatically decides the optimal distribution of the physical memory to partitions and adjusts the memory assignment based on partition load. The administrator reserves physical memory for the shared memory pool, assigns partitions to the pool, and provides access limits to the pool.

Active Memory Sharing can be used to increase memory utilization on the system either by decreasing the global memory requirement or by allowing the creation of additional partitions on an existing system. Active Memory Sharing can be used in parallel with Active Memory Expansion on a system running a mixed workload of several operating system. For example, AIX partitions can take advantage of Active Memory Expansion. Other operating systems take advantage of Active Memory Sharing also.

For additional information regarding Active Memory Sharing, see *PowerVM Virtualization Active Memory Sharing*, REDP-4470.

3.4.7 Active Memory Deduplication

In a virtualized environment, the systems might have a considerable amount of duplicated information that is stored on RAM after each partition has its own operating system, and some of them might even share the same kinds of applications. On heavily loaded systems, this behavior might lead to a shortage of the available memory resources, forcing paging by the Active Memory Sharing partition operating systems, the Active Memory Deduplication pool, or both, which might decrease overall system performance.

Figure 3-13 shows the standard behavior of a system without Active Memory Deduplication enabled on its Active Memory Sharing (shown as AMS in the figure) shared memory pool. Identical pages within the same or different LPARs each require their own unique physical memory page, consuming space with repeated information.

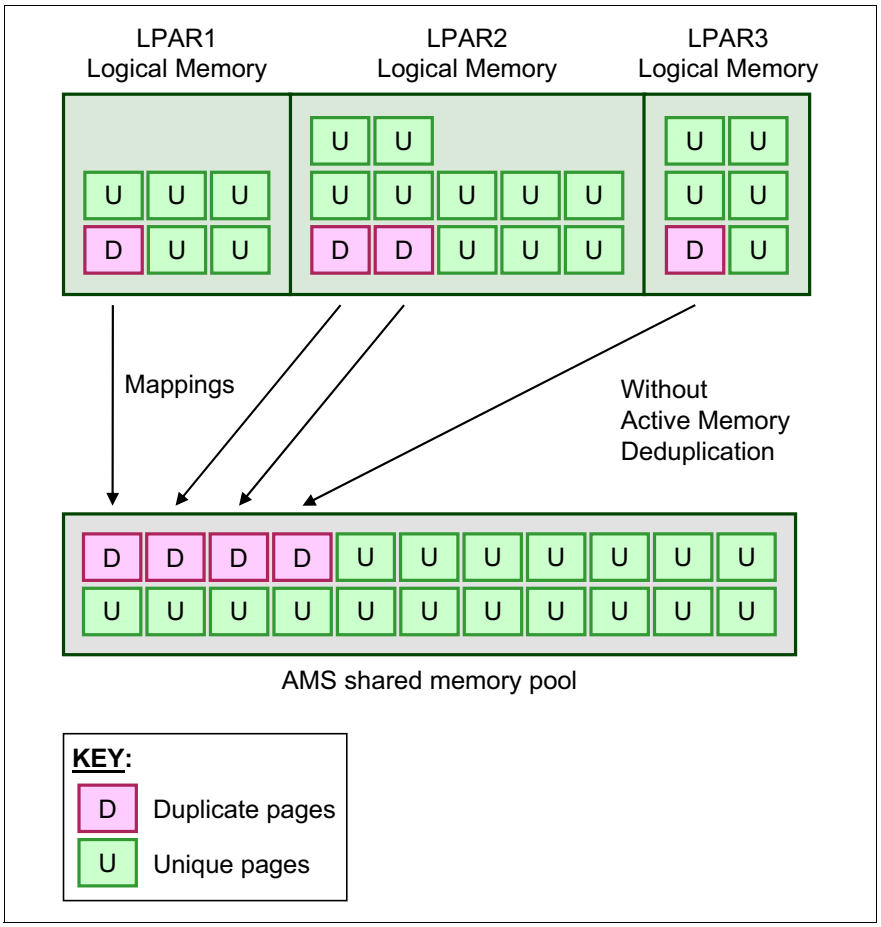


Figure 3-13 Active Memory Sharing shared memory pool without Active Memory Deduplication enabled

Active Memory Deduplication allows the hypervisor to dynamically map identical partition memory pages to a single physical memory page within a shared memory pool. This way enables a better utilization of the Active Memory Sharing shared memory pool, increasing the system's overall performance by avoiding paging. Deduplication can cause the hardware to incur fewer cache misses, which also leads to improved performance.

Figure 3-14 shows the behavior of a system with Active Memory Deduplication enabled on its Active Memory Sharing shared memory pool. Duplicated pages from separate LPARs are stored only once, providing the Active Memory Sharing pool with more free memory.

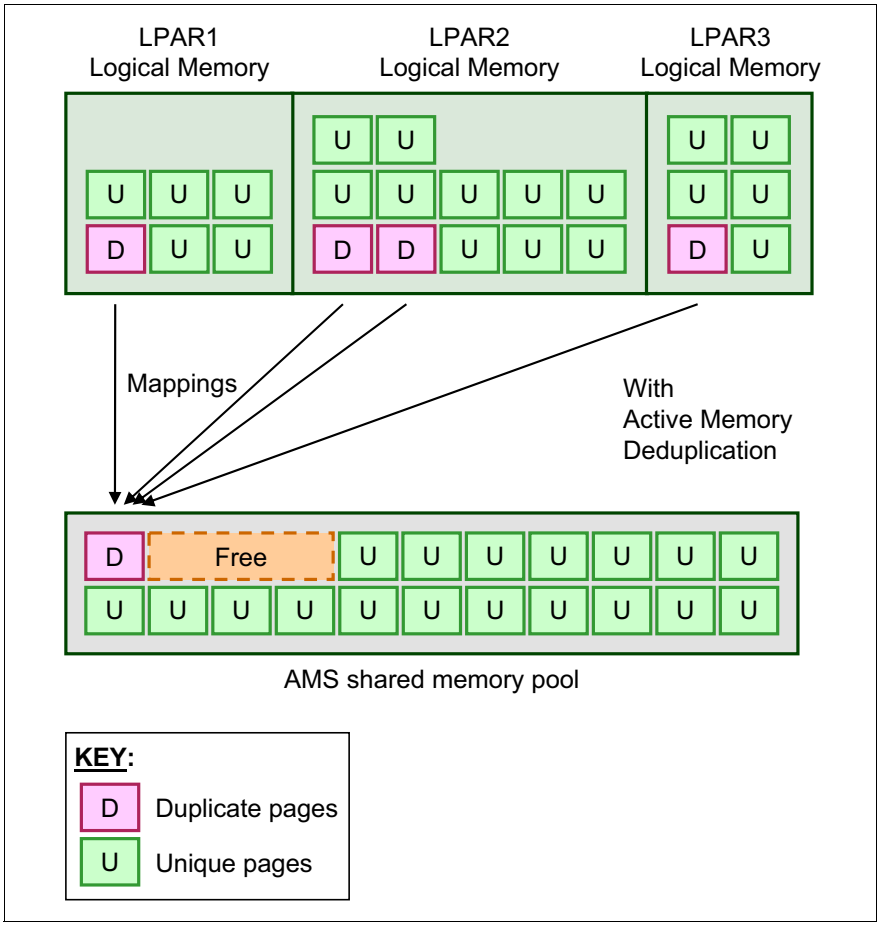


Figure 3-14 Identical memory pages mapped to a single physical memory page with Active Memory Duplication enabled

Active Memory Deduplication depends on the Active Memory Sharing feature to be available, and consumes CPU cycles donated by the Active Memory Sharing pool's Virtual I/O Server (VIOS) partitions to identify deduplicated pages. The operating systems that are running on the Active Memory Sharing partitions can "hint" to the PowerVM Hypervisor that some pages (such as frequently referenced read-only code pages) are particularly good for deduplication.

To perform deduplication, the hypervisor cannot compare every memory page in the Active Memory Sharing pool with every other page. Instead, it computes a small signature for each page that it visits and stores the signatures in an internal table. Each time that a page is inspected, a look-up of its signature is done in the known signatures in the table. If a match is found, the memory pages are compared to be sure that the pages are really duplicates. When a duplicate is found, the hypervisor remaps the partition memory to the existing memory page and returns the duplicate page to the Active Memory Sharing pool.

Figure 3-15 shows two pages being written in the Active Memory Sharing memory pool and having their signatures matched on the deduplication table.

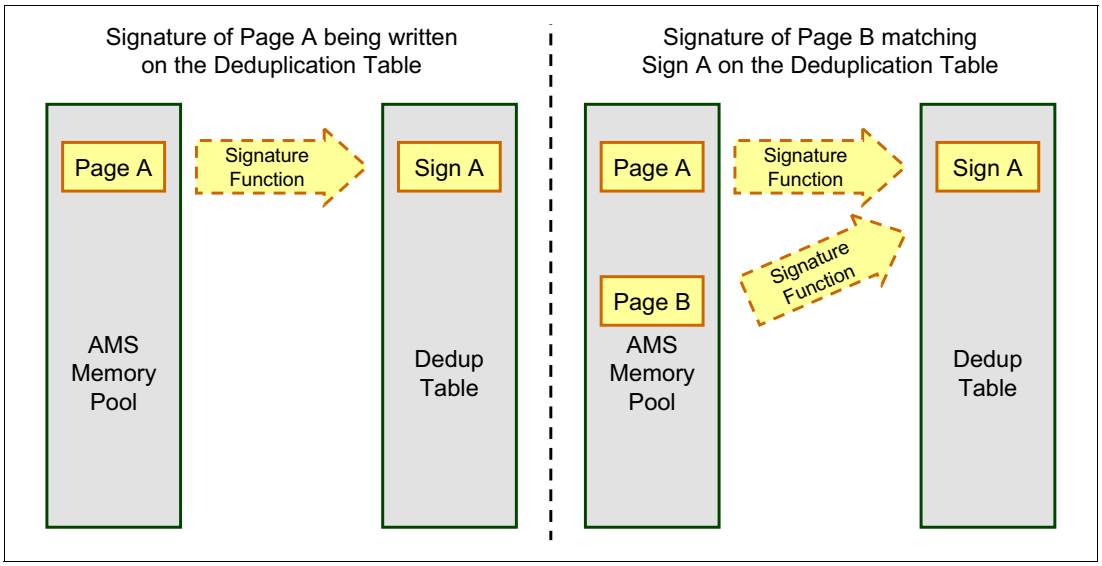


Figure 3-15 Memory pages having their signatures matched by Active Memory Deduplication

From the LPAR perspective, the Active Memory Deduplication feature is completely transparent. If an LPAR attempts to modify a deduplicated page, the hypervisor grabs a free page from the Active Memory Sharing pool, copies the duplicate page contents into the new page, and maps the LPAR's reference to the new page so that the LPAR can modify its own unique page.

System administrators can dynamically configure the size of the deduplication table, ranging from 1/8192 to 1/256 of the configured maximum Active Memory Sharing memory pool size. Having this table be too small might lead to missed deduplication opportunities. Conversely, having a table that is too large might waste a small amount of overhead space.

The management of the Active Memory Deduplication feature is done through a managed console, allowing administrators to take the following steps:

- Enable and disable Active Memory Deduplication at an Active Memory Sharing pool level.
- Display deduplication metrics.
- Display and modify the deduplication table size.

Figure 3-16 shows the Active Memory Deduplication being enabled to a shared memory pool.

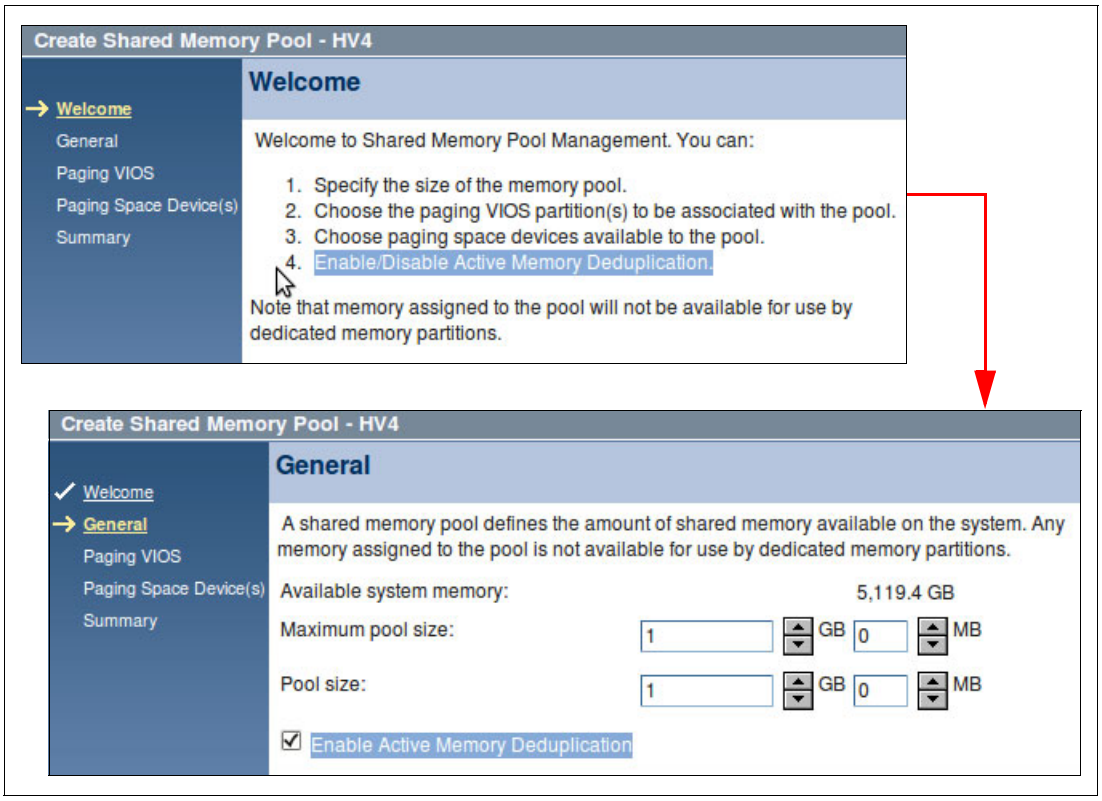


Figure 3-16 Enabling the Active Memory Deduplication for a shared memory pool

The Active Memory Deduplication feature requires the following minimum components:

- PowerVM Enterprise edition
- System firmware level 740
- AIX Version 6: AIX 6.1 TL7 or later
- AIX Version 7: AIX 7.1 TL1 SP1 or later
- IBM i: 7.14 or 7.2 or later
- SLES 11 SP2 or later
- RHEL 6.2 or later

3.4.8 Dynamic Platform Optimizer

Dynamic Platform Optimizer (DPO) is an IBM PowerVM feature that helps the user to configure the logical partition memory and CPU affinity on the POWER7+ processor-based servers, thus, improve performance under some workload scenarios.

On a nonuniform memory access (NUMA) context, the main goal of the DPO is to assign a local memory to the CPUs, thus, reducing the memory access time, because a local memory access is much faster than a remote access.

Accessing remote memory on a NUMA environment is expensive, although common, mainly if the system did a partition migration, or even, if logical partitions are created, suspended, and destroyed frequently, as happens frequently in a cloud environment. In this context, DPO tries to swap remote memory with local memory to the CPU.

Launch Dynamic Platform Optimizer through the HMC command-line interface with the **optmem** command (see Example 3-1). The **lsoptmem** command is able to show important information about current and predicted memory affinity, and also monitor the status of a running optimization process.

Example 3-1 Launching DPO for an LPAR 1

```
#optmem -m <managed_system> -t affinity -o start
```

TIP: While the DPO process is running, the affected LPARs can have up to 20% performance degradation. To explicitly protect partitions from DPO, use the **-x** or **--xid** options of the **optmem** command.

For more information about DPO, see *IBM PowerVM Virtualization Managing and Monitoring*, SG24-7590.

Note: Single-socket systems do not require DPO, and there is no performance penalty when accessing memory in the same card.

3.4.9 Dynamic System Optimizer

Dynamic System Optimizer (DSO) is a PowerVM and AIX feature that autonomously tunes the allocation of system resources to achieve an improvement in system performance. It works by continuously monitoring, through a user space daemon, and analyzing how current workloads impact the system, and then using this information to dynamically reconfigure the system to optimize for current workload requirements. DSO also interacts with POWER7 Performance Monitoring Unit (PMU) to discover the best affinity and page size for the machine workload.

3.4.10 Operating system support for PowerVM

Table 3-5 summarizes the PowerVM features that are supported by the operating systems compatible with the POWER7+ processor-based servers.

Table 3-5 Virtualization features supported by AIX, IBM i and Linux

Feature	AIX 5.3	AIX 6.1	AIX 7.1	IBM i 6.1.1	IBM i 7.1	RHEL 5.8	RHEL 6.3	SLES 10 SP4	SLES 11 SP2
Virtual SCSI	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Virtual Ethernet	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Shared Ethernet Adapter	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Virtual Fibre Channel	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Virtual Tape	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Logical partitioning	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
DLPAR I/O adapter add/remove	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
DLPAR processor add/remove	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Feature	AIX 5.3	AIX 6.1	AIX 7.1	IBM i 6.1.1	IBM i 7.1	RHEL 5.8	RHEL 6.3	SLES 10 SP4	SLES 11 SP2
DLPAR memory add	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
DLPAR memory remove	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes
Micro-Partitioning	Yes	Yes	Yes	Yes	Yes	Yes[a]	Yes[b]	Yes[a]	Yes
Shared dedicated capacity	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Multiple Shared Processor Pools	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Virtual I/O Server	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Integrated Virtualization Manager	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Suspend and resume	No	Yes	Yes	No	Yes[c]	Yes	Yes	No	No
Shared Storage Pools	Yes	Yes	Yes	Yes	Yes[d]	Yes	Yes	Yes	No
Thin provisioning	Yes	Yes	Yes	Yes[e]	Yes[e]	Yes	Yes	Yes	No
Active Memory Sharing	No	Yes	Yes	Yes	Yes	No	Yes	No	Yes
Active Memory Deduplication	No	Yes[f]	Yes[g]	No	Yes[h]	No	Yes	No	Yes
Live Partition Mobility	Yes	Yes	Yes	No	Yes[i]	Yes	Yes	Yes	Yes
Simultaneous multithreading (SMT)	Yes[j]	Yes[k]	Yes	Yes[l]	Yes	Yes[j]	Yes	Yes[j]	Yes
Active Memory Expansion	No	Yes[m]	Yes	No	No	No	No	No	No
Capacity on Demand[n]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
AIX Workload Partitions	No	Yes	Yes	No	No	No	No	No	No

a. This version can only support 10 virtual machines per core.
b. Need RHEL 6.3 Errata upgrade to support 20 virtual machines per core.
c. Requires IBM i 7.1 TR2 with PTF SI39077 or later.
d. Requires IBM i 7.1 TR1.
e. Will become fully provisioned device when used by IBM i.
f. Requires AIX 6.1 TL7 or later.
g. Requires AIX 7.1 TL1 or later.
h. Requires IBM i 7.1.4 or later.
i. Requires IBM i 7.1 TR4 PTF group or later. Access this link for more details: http://bit.ly/11im9sa
j. Only supports two threads.
k. AIX 6.1 up to TL4 SP2 only supports two threads, and supports four threads as of TL4 SP3.
l. IBM i 6.1.1 and up support SMT4.
m. On AIX 6.1 with TL4 SP2 and later.
n. Available on selected models.

3.4.11 Linux support

IBM Linux Technology Center (LTC) contributes to the development of Linux by providing support for IBM hardware in Linux distributions. In particular, the LTC has available tools and code to the Linux communities to take advantage of the POWER7+ technology and develop POWER7+ optimized software.

Table 3-6 lists the support of specific programming features for various versions of Linux.

Table 3-6 Linux support for POWER7 features

Features	Linux releases				Comments
	SLES 10 SP4	SLES 11 SP2	RHEL 5.8	RHEL 6.3	
POWER6 compatibility mode	Yes	Yes	Yes	Yes	-
POWER7 mode	No	Yes	No	Yes	Take advantage of the POWER7+ and POWER7 features.
Strong Access Ordering	No	Yes	No	Yes	Can improve Lx86 performance.
Scale to 256 cores/ 1024 threads	No	Yes	No	Yes	Base OS support is available.
Four-way SMT	No	Yes	No	Yes	Better hardware usage.
VSX support	No	Yes	No	Yes	Full exploitation requires Advance Toolchain.
Distro toolchain mcpu/mtune=p7	No	Yes	No	Yes	SLES11/GA toolchain has minimal P7 enablement necessary to support kernel build.
Advance Toolchain support	Yes, execution restricted to Power6 instructions	Yes	Yes, execution restricted to Power6 instructions	Yes	Alternative GNU Toolchain that explores the technologies available on POWER architecture.
64 KB base page size	No	Yes	Yes	Yes	Better memory utilization, and smaller footprint.
Tickless idle	No	Yes	No	Yes	Improved energy utilization and virtualization of partially to fully idle partitions.

See the following sources for information:

- Advance Toolchain:

  http://ibm.co/1O6nMYI

- Release notes:
  - ftp://linuxpatch.ncsa.uiuc.edu/toolchain/at/at05/suse/SLES_11/release_notes.at05-2.1-0.html
  - ftp://linuxpatch.ncsa.uiuc.edu/toolchain/at/at05/redhat/RHEL5/release_notes.at05-2.1-0.html

3.5 System Planning Tool

The IBM System Planning Tool (SPT) helps you design systems to be partitioned with logical partitions. You can also plan for and design non-partitioned systems by using the SPT. 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 for the following reasons:

- To create reports
- As input to the IBM configuration tool (e-Config)
- To create and deploy partitions on your system (or systems) automatically

System plans that are generated by the SPT can be deployed on the system by the Hardware Management Console (HMC), or Integrated Virtualization Manager (IVM).

Automatically deploy: Ask your IBM representative or IBM Business Partner to use the Customer Specified Placement manufacturing option if you want to automatically deploy your partitioning environment on a new machine. SPT looks for the resource's allocation to be the same as that specified in your `.sysplan` file.

You can create an entirely new system configuration, or you can create a system configuration based on any of the following items:

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

Integration between the System Planning Tool and both the Workload Estimator and IBM Performance Management allows you to create a system that is based on performance and capacity data from an existing system or that is based on new workloads that you specify.

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

Using the SPT is an effective way of documenting and backing up key system settings and partition definitions. With it, the user can create records of systems and export them to their personal workstation or backup system of choice. These same backups can then be imported back onto the same managed console when needed. This step can be useful when cloning systems. enabling the user to import the system plan to any managed console multiple times.

The SPT and its supporting documentation is on the IBM System Planning Tool site:

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

3.6 New PowerVM version 2.2.2 features

Power Systems server coupled with PowerVM technology are designed to help clients build a dynamic infrastructure, reducing costs, managing risk, and improving services levels.

IBM PowerVM V2.2.2 includes VIOS 2.2.2.1-FP26, HMC V7R7.6 and Power Systems firmware level 760, and contains the following enhancements for managing a PowerVM virtualization environment:

- Supports up to 20 partitions per processor, doubling the number of partitions that are supported for each processor. This way provides additional flexibility by reducing the minimum processor entitlement to 5% of a processor.
- Dynamically add virtual I/O adapters to or dynamically remove them from a Virtual I/O Server partition

  HMC V7R7.6 or later automatically runs the add or remove command (**`cfgdev`** or **`rmdev`**) on the Virtual I/O Server for the user. Prior to this enhancement, the user had to manually run these commands on the Virtual I/O Server.
- The user can specify the destination Fibre Channel port for any or all virtual Fibre Channel adapters.
- Virtual I/O Server setup, tuning, and validation is improved by using the Runtime Expert.
- Live Partition Mobility supports up to 16 concurrent LPM activities.
- Shared Storage Pools create pools of storage for virtualized workloads, and can improve storage utilization, simplify administration, and reduce SAN infrastructure costs. The enhancements capabilities enable 16 nodes to partecipate in a Shared Pool configuration, which can improve efficiency, agility, scalability, flexibility, and availability.

  Shared Storage Pools flexibility and availability improvements include the following items:

  - IPv6 and VLAN tagging (IEEE 802.1Q) support for intermodal shared storage pools communication.
  - Cluster reliability and availability improvements.
  - Improved storage utilization statistics and reporting.
  - Nondisruptive rolling upgrades for applying service.
  - Advanced features that accelerate partition deployment, optimize storage utilization, and improve availability through automation.
- New VIOS Performance Advisor analyzes Virtual I/O Server performance, and makes reccomandations for performance optimization.
- PowerVM includes the following new advanced features, enabled by VMControl, that accelerate partition deployment, optimize storage utilization and improve availability through automation:

  - Linked clones allow for sharing of partition images, which greatly accelerates partition deployment and reduces the storage usage.
  - System pool management for IBM workload provides increased flexibility and resource utilization.

  For further details about the appropriate System Director VMControl release, go to the following location:

  `http://www.ibm.com/systems/software/director/vmcontrol`

4

Continuous availability and manageability

This chapter provides information about IBM reliability, availability, and serviceability (RAS) design and features. This set of technologies, implemented on IBM Power Systems servers, improves your architecture's total cost of ownership (TCO) by reducing planned and unplanned down time.

The elements of RAS can be described as follows:

- Reliability: Indicates how infrequently a defect or fault in a server occurs
- Availability: Indicates how infrequently the functionality of a system or application is impacted by a fault or defect
- Serviceability: Indicates how well faults and their effects are communicated to system managers and how efficiently and nondisruptively the faults are repaired

Each successive generation of IBM servers is designed to be more reliable than the previous server family. POWER7+ processor-based servers have new features to support new levels of virtualization, help ease administrative burden, and increase system utilization.

Reliability starts with components, devices, and subsystems designed to be fault-tolerant. POWER7+ uses lower voltage technology, improving reliability with stacked latches to reduce soft error susceptibility. During the design and development process, subsystems go through rigorous verification and integration testing processes. During system manufacturing, systems go through a thorough testing process to help ensure high product quality levels.

The processor and memory subsystem contain features that are designed to avoid or correct environmentally induced, single-bit, intermittent failures. The features can also handle solid faults in components, including selective redundancy to tolerate certain faults without requiring an outage or parts replacement.

© Copyright IBM Corp. 2013. All rights reserved.

4.1 Reliability

Highly reliable systems are built with highly reliable components. On IBM POWER 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 that are designed with fewer components and interconnects have fewer opportunities to fail. Simple design choices such as integrating processor cores on a single POWER chip can dramatically reduce the opportunity for system failures. In this case, an 8-core server can include one quarter as many processor chips (and chip socket interfaces) as with a dual core processor design. Not only does this case reduce the total number of system components, it reduces the total amount of heat that is generated in the design, resulting in an additional reduction in required power and cooling components. POWER7+ processor-based servers also integrate L3 cache into the processor chip for a higher integration of parts.

Parts selection also plays a critical role in overall system reliability. IBM uses three grades of components with grade 3 being defined as industry standard (“off-the-shelf” components). 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.

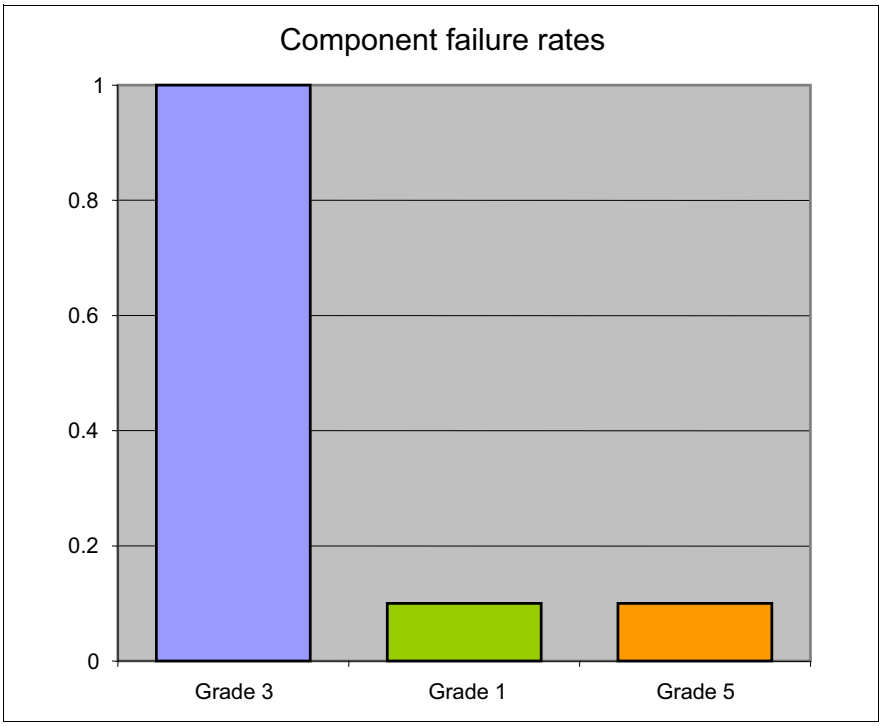


Figure 4-1 Component failure rates

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. All POWER processor-based systems are carefully packaged to ensure adequate cooling. Critical system components such as the POWER7+ processor chips are positioned on the planar so that they receive clear air flow during operation. In addition, POWER 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 (those that most affect system availability*)* are protected with redundancy and the ability to be repaired concurrently.

The use of these redundant components allows the system to remain operational:

- POWER7+ cores, which include redundant bits in L1 instruction and data caches, L2 caches, and L2 and L3 directories
- Power 710 and Power 730 main memory DIMMs, which use an innovative ECC algorithm, from IBM research, that improves bit-error correction and memory failures
- Redundant and hot-swap cooling
- Redundant and hot-swap power supplies

For maximum availability, be sure 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. Tower form factor power cords must be plugged into two independent power sources to achieve maximum availability.

Before ordering: Check your configuration for optional redundant components before ordering your system.

4.2 Availability

First-failure data capture (FFDC) is the capability of IBM hardware and microcode to continuously monitor hardware functions. This process includes predictive failure analysis, which is the ability to track intermittent correctable errors and to take components offline before they reach the point of hard failure. This way avoids causing a system outage.

The POWER7+ family of systems can do the following automatic functions:

- Self-diagnose and self-correct errors during run time.
- Automatically reconfigure to mitigate potential problems from suspect hardware.
- Self-heal or automatically substitute good components for failing components.

Remember: Error detection and fault isolation is independent of the operating system in POWER7+ processor-based servers.

This chapter describes IBM POWER7+ processor-based systems technologies, focused on keeping a system running. For a specific set of functions that are focused on detecting errors before they become serious enough to stop computing work, see 4.3.1, “Detecting” on page 149.

4.2.1 Partition availability priority

POWER7+ systems can assign availability priorities to partitions. If the system detects that a processor core is about to fail, it is taken offline. If the partitions on the system require more processor units than remain in the system, the firmware determines which partition has the lowest priority and attempts to claim the needed resource. On a properly configured POWER processor-based server, this capability allows the system manager to ensure that capacity is first obtained from a low-priority partition instead of a high-priority partition.

This capability gives the system an additional stage before an unplanned outage. If insufficient resources exist to maintain full system availability, the server attempts to maintain partition availability according to user-defined priority.

Partition availability priority is assigned to partitions by using a *weight value* or integer rating. The lowest priority partition is rated at 0 (zero) and the highest priority partition is rated at 255. The default value is set to 127 for standard partitions and 192 for Virtual I/O Server (VIOS) partitions. You can vary the priority of individual partitions through the hardware management console.

4.2.2 General detection and deallocation of failing components

Runtime correctable or recoverable errors are monitored to determine whether 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 POWER 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, 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 way prevents the faulty hardware from affecting system operation again; the repair action is deferred to a more convenient, less critical time.

The following components have the capability to be persistently deallocated:

- Processor
- L2 and L3 cache lines (Cache lines are dynamically deleted.)
- Memory
- Deconfigure or bypass failing I/O adapters

Processor instruction retry

As introduced with the POWER6 technology, the POWER7+ processor can retry processor instructions and do alternate processor recovery for several core-related faults. In this way, exposure to both permanent and intermittent errors in the processor core are significantly reduced.

Intermittent errors are generally not repeatable, often because of cosmic rays or other sources of radiation.

With the instruction retry function, when an error is encountered in the core, in caches and certain logic functions, the POWER7+ processor first automatically retries the instruction. If the source of the error was truly transient, the instruction succeeds and the system can continue as before.

Before POWER6: On IBM systems prior to POWER6, such an error typically caused a checkstop

Alternate processor retry

Hard failures are more difficult; they are permanent errors that are replicated each time that the instruction is repeated. Retrying the instruction does not help in this situation because the instruction will continue to fail.

As introduced with POWER6, POWER7+ processors can extract the failing instruction from the faulty core and retry it elsewhere in the system. The failing core is then dynamically deconfigured and scheduled for replacement.

Dynamic processor deallocation

Dynamic processor deallocation enables automatic deconfiguration of processor cores when patterns of recoverable core-related faults 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. The POWER Hypervisor then dynamically deconfigures the failing core and notifies the system administrator that a replacement is needed. The entire process is transparent to the partition owning the failing instruction.

Single processor checkstop

As in the POWER6 processor, the POWER7+ processor provides single core check-stopping for certain processor logic, command, or control errors that cannot be handled by the availability enhancements in the preceding section.

This approach significantly reduces the probability of any one processor affecting total system availability by containing most processor checkstops to the partition that was using the processor at the time that full checkstop goes into effect.

Even with all these availability enhancements to prevent processor errors from affecting system-wide availability, errors might occur that can result in a system-wide outage.

4.2.3 Memory protection

A memory protection architecture that provides good error resilience for a relatively small L1 cache might be inadequate for protecting the much larger system main store. Therefore, a variety of protection methods are used in all POWER processor-based systems to avoid uncorrectable errors in memory.

Memory protection plans must account for many factors, including the following factors:

- Size
- Desired performance
- Memory array manufacturing characteristics

POWER7+ processor-based systems have various protection schemes designed to prevent, protect, or limit the effect of errors in main memory:

- Chipkill

  Chipkill is an enhancement that enables a system to sustain the failure of an entire DRAM chip. An ECC word uses 18 DRAM chips from two DIMM pairs, and a failure on any of the DRAM chips can be fully recovered by the ECC algorithm. The system can continue indefinitely in this state with no performance degradation until the failed DIMM can be replaced.

- 72-byte ECC

  In POWER7+, an ECC word consists of 72 bytes of data. Of these, 64 bytes are used to hold application data. The remaining eight bytes are used to hold check bits and additional information about the ECC word.

  This innovative ECC algorithm from IBM research works on DIMM pairs on a rank basis. (A *rank* is a group of nine DRAM chips.) With this ECC code, the system can dynamically recover from an entire DRAM failure (by Chipkill) but can also correct an error even if another *symbol* (a byte, accessed by a 2-bit line pair) experiences a fault (an improvement from the double error detection or single error correction ECC implementation found on the POWER6 processor-based systems).

- Hardware scrubbing

  Hardware scrubbing is a method used to handle intermittent errors. IBM POWER processor-based systems periodically address all memory locations. Any memory locations with a correctable error are rewritten with the correct data.

- Cyclic redundancy check (CRC)

  The bus that is transferring data between the processor and the memory uses CRC error detection with a failed operation-retry mechanism and the ability to dynamically retune the bus parameters when a fault occurs. In addition, the memory bus has spare capacity to substitute a data bit-line whenever it is determined to be faulty.

POWER7+ memory subsystem

The POWER7+ processor chip contains two memory controllers with four channels per memory controller. Each channel connects to a single DIMM, but as the channels work in pairs, a processor chip can address four DIMM pairs, two pairs per memory controller.

The bus transferring data between the processor and the memory uses CRC error detection with a failed operation retry mechanism and the ability to dynamically retune bus parameters when a fault occurs. In addition, the memory bus has spare capacity to substitute a spare data bit-line for one that is determined to be faulty.

Advanced memory buffer chips are exclusive to IBM and help to increase performance, acting as read/write buffers. The Power 710 and the Power 730 use one memory controller. Advanced memory buffer chips are on the memory cards and support two DIMMs each.

Memory page deallocation

Although coincident cell errors in separate memory chips are statistically rare, IBM POWER7+ processor-based systems can contain these errors by using a memory page deallocation scheme for partitions that are running IBM AIX and IBM i operating systems, and also for memory pages that are 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.

Pages that are used by the POWER Hypervisor are deallocated as soon as the page is released. In other cases, the POWER Hypervisor notifies the owning partition that the page must be deallocated. Where possible, the operating system moves any data currently contained in that memory area to another memory area and removes the pages associated with this error from its memory map, no longer addressing these pages. The operating system performs memory page deallocation without any user intervention and is transparent to users and applications.

The POWER Hypervisor maintains a list of pages marked for deallocation during the current platform initial program load (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 when memory pages are included that need to be deallocated.

Finally, if an uncorrectable error in memory is discovered, the logical memory block that is associated with the address that has the uncorrectable error is marked for deallocation by the POWER Hypervisor. This deallocation becomes effective on a partition reboot if the logical memory block is assigned to an active partition at the time of the fault.

In addition, the system will deallocate the entire memory group that is associated with the error on all subsequent system reboots until the memory is repaired. This precaution is intended to guard against future uncorrectable errors while waiting for parts replacement.

Memory persistent deallocation

Defective memory that is discovered at boot time is automatically switched off. If the service processor detects a memory fault at boot time, it marks the affected memory as bad so that it is not used on subsequent reboots.

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. Because these activities reduce the amount of CoD memory that is available for future use, repair of the faulty memory must be scheduled as soon as 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.

Depending on the configuration of the system, the HMC IBM Service Focal Point™, OS Service Focal Point, or service processor will receive a notification of the failed component, and will trigger a service call.

4.2.4 Cache protection

POWER7+ processor-based systems are designed with cache protection mechanisms, including cache-line delete in both L2 and L3 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, and L2 and L3 directories.

L1 instruction and data array protection

The POWER7+ processor instruction and data caches are protected against intermittent errors using processor instruction retry and against permanent errors by alternate processor recovery, both mentioned previously. L1 cache is divided into sets. POWER7+ processor can deallocate all but one before doing a processor instruction retry.

In addition, faults in the Segment Lookaside Buffer (SLB) array are recoverable by the POWER Hypervisor. The SLB is used in the core to do address translation calculations.

L2 and L3 array protection

The L2 and L3 caches in the POWER7+ processor are protected with double-bit detect single-bit correct error detection code (ECC). Single-bit errors are corrected before forwarding to the processor and are subsequently written back to the L2 and L3 cache.

In addition, the caches maintain a cache-line delete capability. A threshold of correctable errors that are detected on a cache line can result in the data in the cache line being purged and the cache line removed from further operation without requiring a reboot. An ECC uncorrectable error detected in the cache can also trigger a purge and delete of the cache line. This results in no loss of operation because an unmodified copy of the data can be held on system memory to reload the cache line from main memory. Modified data is handled through Special Uncorrectable Error handling.

L2 and L3 deleted cache lines are marked for persistent deconfiguration on subsequent system reboots until they can be replaced.

4.2.5 Special Uncorrectable Error handling

Although rare, an uncorrectable data error can occur in memory or cache. IBM POWER processor-based systems attempt to limit the impact of an uncorrectable error to the least possible disruption, 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, as in the following 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.
- The L2 and L3 cache of the POWER7+ processor-based systems can hold an unmodified copy of data in a portion of main memory. In this case, an uncorrectable error simply triggers a reload of a cache line from main memory.

In cases where the data cannot be recovered from another source, a technique named Special Uncorrectable Error (SUE) handling is used to prevent an uncorrectable error in memory or cache from immediately causing the system to terminate. That is, the system tags the data and determines whether it will ever be used again:

- If the error is irrelevant, SUE will not force a checkstop.
- If data is used, termination can be limited to the program/kernel or hypervisor that owns the data, or freeze the I/O adapters that are controlled by an I/O hub controller if data is going to be transferred to an I/O device.

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 5.2, or later, or Linux, and a process attempts to use the data, the operating system is informed of the error and might terminate, or only terminate a specific process that is associated with the corrupt data, depending on the operating system and firmware level and whether the data was associated with a kernel or non-kernel process.

Only in the case where the corrupt data is used by the POWER Hypervisor must the entire system be rebooted, thereby preserving overall system integrity.

Depending on system configuration and the source of the data, errors encountered during I/O operations might 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 from 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 might include the entire PHB chip, or simply a single bridge, resulting in the loss of all I/O operations that use the frozen hardware until a power-on reset of the PHB is done. The impact to partitions depends on how the I/O is configured for redundancy. In a server configured for failover availability, redundant adapters spanning multiple PHB chips can enable the system to recover transparently, without partition loss.

4.2.6 PCI Enhanced Error Handling

IBM estimates that PCI adapters can account for a significant portion of the hardware-based errors on a large server. Although 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 certain cases, they might 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 certain cases, an error in the adapter might 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.

PCI Enhanced Error Handling (EEH) enabled adapters respond to a special data packet that is generated from the affected PCI slot hardware by calling system firmware, which will examine the affected bus, allow the device driver to reset it, and continue without a system

reboot. For Linux, EEH support extends to the majority of frequently used devices, although various third-party PCI devices might not provide native EEH support.

To detect and correct PCIe bus errors, POWER7+ processor-based systems use CRC detection and instruction-retry correction; for PCI-X, the systems use ECC.

Figure 4-2 shows the location and various mechanisms that are used throughout the I/O subsystem for PCI Enhanced Error Handling.

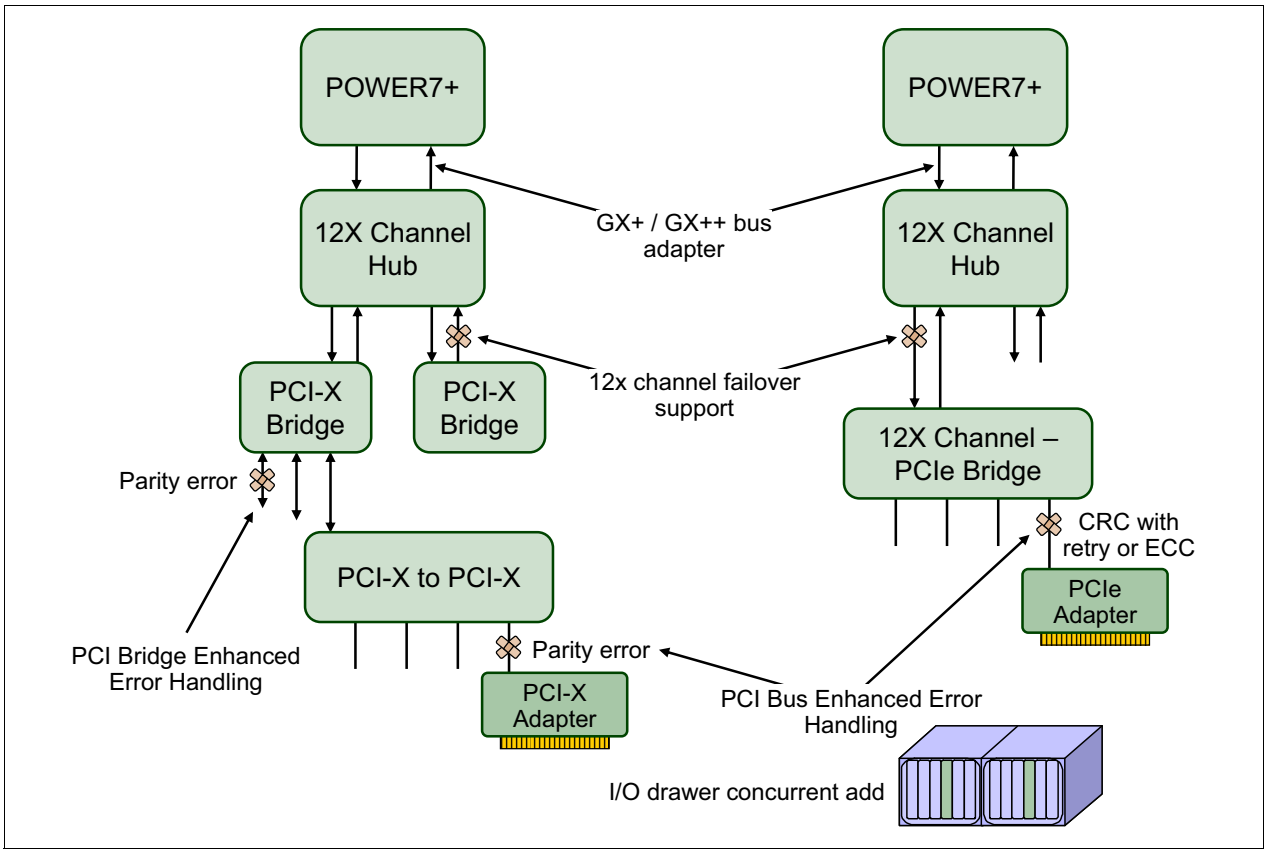


Figure 4-2 PCI Enhanced Error Handling

4.3 Serviceability

IBM Power Systems design considers both IBM and the client's needs. The IBM Serviceability Team, enhanced the base service capabilities and continues to implement a strategy that incorporates best-of-its-kind service characteristics from diverse IBM Systems offerings.

The purpose of serviceability is to repair the system while attempting to minimize or eliminate service cost (within budget objectives), while maintaining high customer satisfaction. Serviceability includes system installation, MES (system upgrades/downgrades), and system maintenance/repair. Depending on the system and warranty contract, service may be performed by the customer, an IBM representative, or an authorized warranty service provider.

The serviceability features that are delivered in this system provide a highly efficient service environment by incorporating the following attributes:

- Design for customer setup (CSU), customer installed features (CIF), and customer-replaceable units (CRU)
- Error detection and fault isolation (ED/FI)
- First-failure data capture (FFDC)
- Converged service approach across multiple IBM server platforms

By delivering on these goals, IBM Power Systems servers enable faster and more accurate repair, and reduce the possibility of human error.

Client control of the service environment extends to firmware maintenance on all of the POWER processor-based systems. 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, notifying, and repairing found in all POWER processor-based systems.

4.3.1 Detecting

The first and most crucial component of a solid serviceability strategy is the ability to accurately and effectively detect errors when they occur. Although not all errors are a guaranteed threat to system availability, those that go undetected can cause problems because the system has no opportunity to evaluate and act if necessary. Power processor-based systems employ IBM System z® server-inspired error detection mechanisms, extending from processor cores and memory to power supplies and hard drives.

Service processor

The service processor is a microprocessor that is powered separately from the main instruction processing complex. The service processor provides the capabilities for the following items:

- POWER Hypervisor (system firmware) and Hardware Management Console connection surveillance
- Several remote power control options
- Reset and boot features
- 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. By using an architected operating system interface, the service processor notifies the operating system of potential environmentally related problems 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 in the following circumstances:

- The operating temperature exceeds the critical level (for example, failure of air conditioning or air circulation around the system).
- The system fan speed is out of operational specification (for example, because of multiple fan failures).
- The server input voltages are out of operational specification.

The service processor can immediately shut down a system in the following circumstances:

- Temperature exceeds the critical level or remains above the warning level for too long.
- Internal component temperatures reach critical levels.
- Non-redundant fan failures occur.

The service processor provides the following features:

- Placing calls

  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.

- Mutual surveillance

  The service processor monitors the operation of the firmware during the boot process, and also monitors the hypervisor for termination. The hypervisor monitors the service processor and can perform a reset and reload if it detects the loss of the service processor. If the reset/reload operation does not correct the problem with the service processor, the hypervisor notifies the operating system; the operating system can then take appropriate action, including calling for service.

- Availability

  The POWER7+ family of systems continues to offer and introduce significant enhancements designed to increase system availability.

  As in POWER6, POWER6+, and POWER7, the POWER7+ processor has the ability to do processor instruction retry and alternate processor recovery for several core-related faults. This ability significantly reduces exposure to both hard (logic) and soft (transient) errors in the processor core. Soft failures in the processor core are transient (intermittent) errors, often because of cosmic rays or other sources of radiation, and generally are not repeatable. When an error is encountered in the core, the POWER7+ processor first automatically retries the instruction. If the source of the error was truly transient, the instruction succeeds and the system continues as before. On IBM systems before POWER6, this error caused a checkstop.

  Hard failures are more difficult; they are true logical errors that are replicated each time the instruction is repeated. Retrying the instruction does not help in this situation. As in POWER6, POWER6+, and POWER7, all POWER7+ processors can extract the failing instruction from the faulty core and retry it elsewhere in the system for several faults, after which the failing core is dynamically deconfigured and called out for replacement. These systems are designed to avoid a full system outage.

- Uncorrectable error recovery

  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.

  The auto-restart (reboot) option must be enabled from the Advanced System Management Interface (ASMI) or from the Control (Operator) Panel.

Figure 4-3 shows this option using the ASMI.

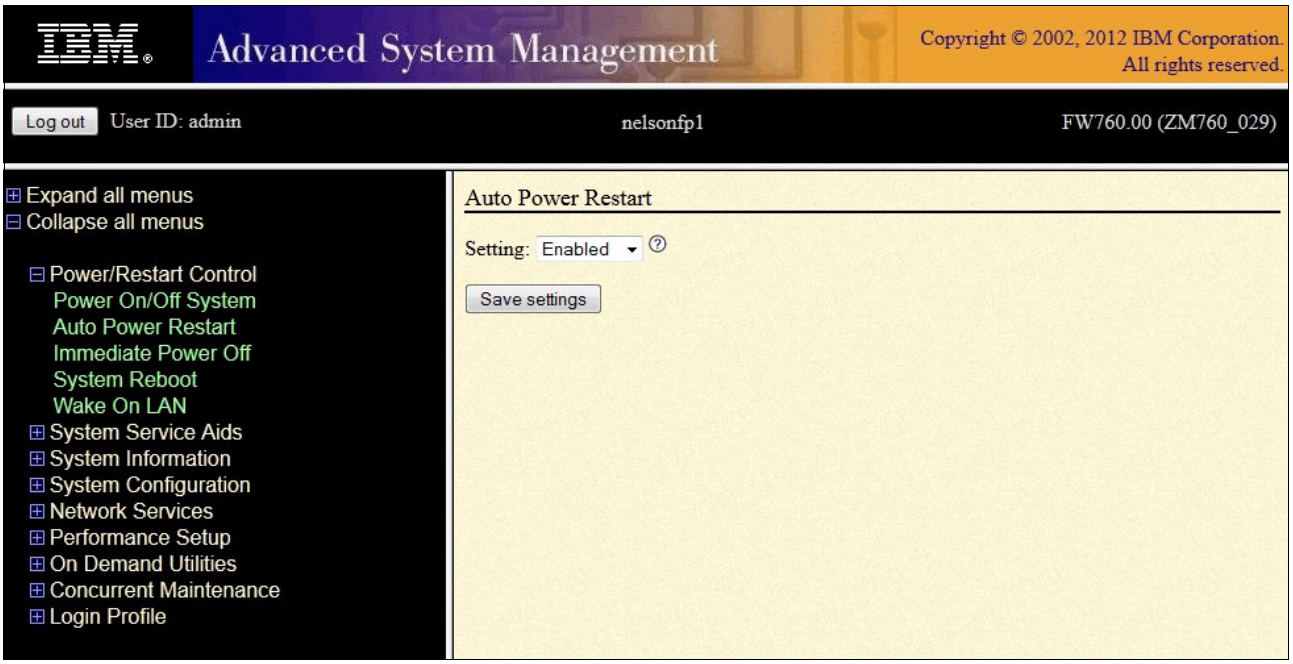


Figure 4-3 ASMI Auto Power Restart setting panel

- Partition availability priority

  Availability priorities can be assigned to partitions. If an alternate processor recovery event requires spare processor resources to protect a workload, when no other means of obtaining the spare resources is available, the system determines which partition has the lowest priority and attempts to claim the needed resource. On a properly configured POWER7+ processor-based server, this way allows that capacity to be first obtained from, for example, a test partition instead of a financial accounting system.

- POWER7+ cache availability

  The L2 and L3 caches in the POWER7+ processor are protected with double-bit detect, single-bit correct error detection code (ECC). In addition, the caches maintain a cache line delete capability. A threshold of correctable errors detected on a cache line can result in the data in the cache line being purged and the cache line removed from further operation without requiring a reboot. An ECC uncorrectable error detected in the cache can also trigger a purge and delete operation of the cache line. This step results in no loss of operation if the cache line contained data that is unmodified from what was stored in system memory. Modified data would be handled through Special Uncorrectable Error handling. L1 data and instruction caches also have a retry capability for intermittent error and a cache set delete mechanism for handling solid failures. In addition, the POWER7+ processors also have the ability to dynamically substitute a faulty bit-line in an L3 cache dedicated to a processor with a spare bit-line.

- Fault monitoring

  Built-in self-test (BIST) checks processor, cache, memory, and associated hardware that is required for proper booting of the operating system, when the system is powered on at the initial installation 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 to 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.

- Concurrent access to the service processors menus of the ASMI

  This access allows nondisruptive abilities to change system default parameters, interrogate service processor progress and error logs, and set and reset server indicators (Guiding Light for midrange and high-end servers, Light Path for low-end servers), accessing all service processor functions without having to power down the system to the standby state. This allows the administrator or service representative to dynamically access the menus from any web browser-enabled console that is attached to the Ethernet service network, concurrently with normal system operation.

- Managing the interfaces for connecting uninterruptible power source systems to the POWER processor-based systems, performing timed power-on (TPO) sequences, and interfacing with the power and cooling subsystem

Error checkers

IBM POWER 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:

- Continuous monitoring of system operations to detect potential calculation errors.
- Attempts to isolate physical faults based on runtime detection of each unique failure.
- Ability to initiate a wide variety of recovery mechanisms designed to correct the problem. The POWER 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 logic circuitry is used to limit the domain of an error to the first checker that encounters the error. In this way, runtime error diagnostics can be deterministic so 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

First-failure data capture (FFDC) is an error isolation technique. It 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 re-create 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 is 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 from the associated logic. In firmware, this data consists of return codes, function calls, and so forth.

FFDC *check stations* are carefully positioned within server logic and data paths to ensure that potential errors can be quickly identified and accurately tracked to a FRU.

This proactive diagnostic strategy is a significant improvement over the classic, less accurate *reboot and diagnose* service approaches.

Figure 4-4 shows a schematic of a fault isolation register implementation.

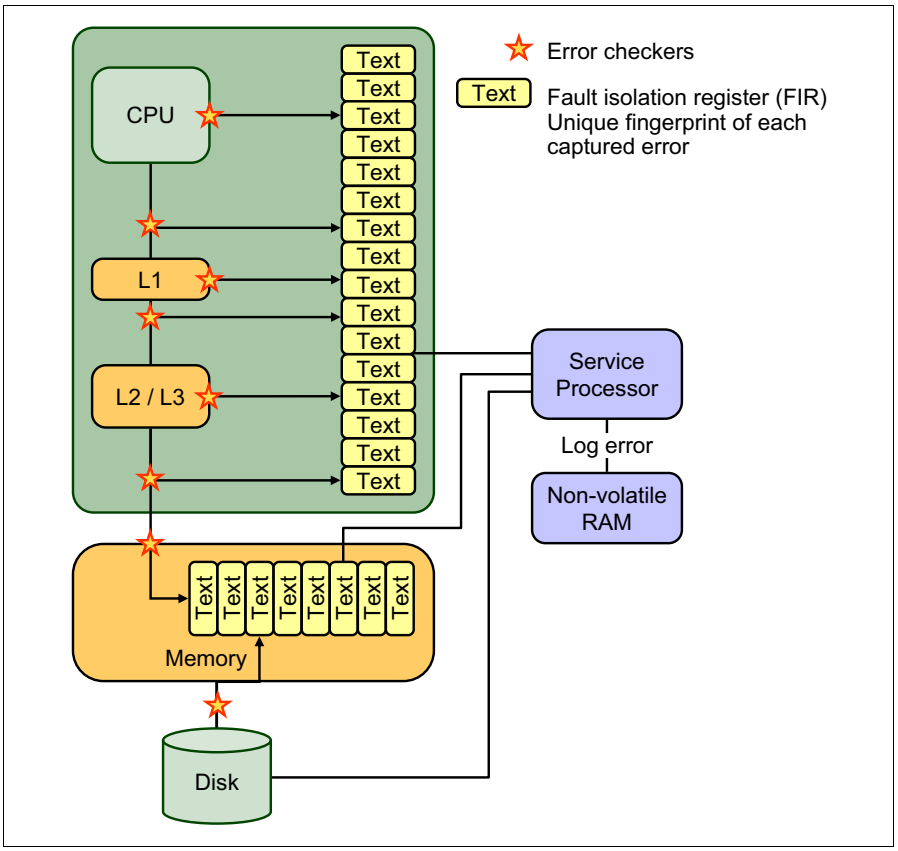


Figure 4-4 Schematic of FIR implementation

Fault isolation

The service processor interprets error data that is captured by the FFDC checkers (saved in the FIRs or other firmware-related data capture methods) 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 might indicate that a service action point was reached, where the event exceeded a predetermined threshold or was unrecoverable. Based on the isolation analysis, recoverable error-threshold counts can be incremented. No specific service action is necessary when the event is recoverable.

When the event requires a service action, additional required information is collected to service the fault. For unrecoverable errors or for recoverable events that meet or exceed their service threshold (meaning that a service action point was reached) a request for service is initiated through an error logging component.

4.3.2 Diagnosing

General diagnostic objectives are to detect and identify problems so that they can be resolved quickly. IBM diagnostics strategy includes the following elements:

- Provide a common error code format equivalent to a system reference code, system reference number, checkpoint, or firmware error code.
- Provide fault detection and problem isolation procedures. Support remote connection ability to be used by the IBM Remote Support Center or IBM Designated Service.
- Provide interactive intelligence within the diagnostics with detailed online failure information while connected to IBM back-end system.

Using the extensive network of advanced and complementary error detection logic that is built directly into hardware, firmware, and operating systems, the IBM Power Systems servers can perform considerable self-diagnosis.

Because of the FFDC technology that is designed into IBM servers, re-creating diagnostics for failures or requiring user intervention is not necessary. Solid and intermittent errors are designed to be correctly detected and isolated at the time that the failure occurs. Runtime and boottime diagnostics fall into this category.

Boot time

When an IBM Power Systems server powers up, the service processor initializes the system hardware. Boot-time diagnostic testing uses a multitier approach for system validation, starting with managed low-level diagnostics that are supplemented with system firmware initialization and configuration of I/O hardware, followed by OS-initiated software test routines. Boot-time diagnostic routines include the following items:

- 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 or not the system processors are operational. Boot-time BISTs can 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 determines which of the diagnostics are required to be started to ensure correct operation, based on the way that the system was powered off, or on the boot-time selection menu.

Run time

All Power Systems servers can monitor critical system components during run time, and they can take corrective actions when recoverable faults occur. IBM 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 runtime diagnostic capabilities originate with the service processor. Extensive diagnostic and fault analysis routines were 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 by using logic derived from IBM engineering expertise to count recoverable errors (called thresholding) and

predict when corrective actions must be automatically initiated by the system. These actions can include the following items:

- Requests for a part to be replaced
- Dynamic 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 often works 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

In the unlikely event that a system hardware or environmentally induced failure is diagnosed, IBM Power Systems servers report the error through various mechanisms. The analysis result is stored in system NVRAM. Error log analysis (ELA) can be used to display the failure cause and the physical location of the failing hardware.

With the integrated service processor, the system can automatically send an alert through a phone line to a pager, or call for service in the event of a critical system failure. A hardware fault also illuminates the amber system fault LED, located on the system unit, to alert the user of an internal hardware problem.

On POWER7+ processor-based servers, hardware and software failures are recorded in the system log. When a management console is attached, an ELA routine analyzes the error, forwards the event to the Service Focal Point (SFP) application running on the management console, and has the capability to notify the system administrator that it has isolated a likely cause of the system problem. The service processor event log also records unrecoverable checkstop conditions, forwards them to the SFP application, and notifies the system administrator. After the information is logged in the SFP application, if the system is properly configured, a call-home service request is initiated and the pertinent failure data with service parts information and part locations is sent to the IBM service organization.This information will also contain the client contact information as defined in the IBM Electronic Service Agent (ESA) guided setup wizard.

Error logging and analysis

When the root cause of an error is identified by a fault isolation component, an error log entry is created with basic data such as the following examples:

- An error code that uniquely describes the error event
- The location of the failing component
- The part number of the component to be replaced, including pertinent data such as engineering and manufacturing levels
- Return codes
- Resource identifiers
- FFDC data

Data that contains information about the effect that the repair will have on the system is also included. Error log routines in the operating system and FSP can then use this information

and decide whether the fault is a call-home candidate. If the fault requires support intervention, a call is placed with service and support, and a notification is sent to the contact that is defined in the ESA-guided setup wizard.

Remote support

The Remote Management and Control (RMC) subsystem is delivered as part of the base operating system, including the operating system that runs 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 several other functions also, but these are not used for the service infrastructure.

Service Focal Point (SFP)

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 be reported only once, regardless of how many logical partitions experience the potential effect of the error. The Manage Serviceable Events task on the management console 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 to relay error information to the Hardware Management Console. For global events (platform unrecoverable errors, for example), the service processor also forwards error notification of these events to the Hardware Management Console, providing a redundant error-reporting path in case of errors in the Remote Management and Control subsystem network.

The first occurrence of each failure type is recorded in the Manage Serviceable Events task on the management console. This task then filters and maintains a history of duplicate reports from other logical partitions on the service processor. It then looks at 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

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 that is collected depends on the invocation method but includes information such as 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 either to assist the service support organization with preparing a service action plan for the service representative or for additional analysis.

System-dump handling

In certain circumstances, an error might require a dump to be automatically or manually created. In this event, it is off-loaded to the management console. Specific management console information is included as part of the information that can optionally be sent to IBM support for analysis. If additional information that relates to the dump is required, or if viewing the dump remotely becomes necessary, the management console dump record notifies the IBM support center regarding on which management console the dump is located.

4.3.4 Notifying

After a Power Systems server detects, diagnoses, and reports an error to an appropriate aggregation point, it then takes steps to notify the client, and if necessary the IBM support organization. Depending on the assessed severity of the error and support agreement, this client notification might range from a simple notification to having field service personnel automatically dispatched to the client site with the correct replacement part.

Client 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 *Client Notify*. Clients 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. These events include the following examples:

- Network events such as 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 (although these events do not necessarily require a part replacement or repair action)

Client Notify events are serviceable events, by definition, because they indicate that something happened that requires client awareness if the client wants to take further action. These events can always be reported back to IBM at the discretion of the client.

Call home

Call home refers to an automatic or manual call from a customer location to an IBM support structure with error log data, server status, or other service-related information. The call home feature invokes the service organization so that the appropriate service action can begin. Call home can be done through HMC or most non-HMC managed systems. Although configuring a call home function is optional, clients are encouraged to implement this feature to obtain service enhancements such as reduced problem determination and faster and potentially more accurate transmittal of error information. In general, using the call home feature can result in increased system availability. The Electronic Service Agent application can be configured for automated call home. See 4.4.4, “Electronic Services and Electronic Service Agent” on page 169 for specific details.

Vital product data and inventory management

Power Systems store vital product data (VPD) internally, which keeps a record of how much memory is installed, how many processors are installed, the manufacturing level of the parts, and so on. These records provide valuable information that can be used by remote support and service representatives, enabling the representatives to provide assistance in keeping the firmware and software current on the server.

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 that is related to the error, along with any service actions taken by the service representative, is 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 servicing

The final component of a comprehensive design for serviceability is the ability to effectively locate and replace parts requiring service. POWER processor-based systems use a combination of visual cues and guided maintenance procedures to ensure that the identified part is replaced correctly, every time.

Packaging for service

The following service enhancements are included in the physical packaging of the systems to facilitate service:

- Color coding (touch points)
  - Terra-cotta-colored touch points indicate that a component (FRU or CRU) can be concurrently maintained.
  - Blue-colored touch points delineate components that are not concurrently maintained (those that require the system to be turned off for removal or repair).
- Tool-less design

  Selected IBM systems support tool-less or simple tool designs. These designs require no tools, or require basic tools such as flathead screw drivers, to service the hardware components.
- Positive retention

  Positive retention mechanisms help to ensure proper connections between hardware components, such as from cables to connectors, and between two cards that attach to each other. Without positive retention, hardware components risk becoming loose during shipping or installation, preventing a good electrical connection. Positive retention mechanisms such as latches, levers, thumb-screws, pop Nylatches (U-clips), and cables are included to help prevent loose connections and aid in installing (seating) parts correctly. These positive retention items do not require tools.

Light Path

The Light Path LED feature is for low-end systems, including Power Systems through models 710 and 730, that can be repaired by clients. In the Light Path LED implementation, when a fault condition is detected on the POWER7 or POWER7+ processor-based system, an amber FRU fault LED is illuminated, which is then rolled up to the system fault LED. The Light Path system pinpoints the exact part by lighting the amber FRU fault LED that is associated with the part to be replaced.

The system can clearly identify components for replacement by using specific component level LEDs, and can also guide the servicer directly to the component by signaling (remaining on, or *solid*) the system fault LED, enclosure fault LED, and the component FRU fault LED.

After the repair, the LEDs shut off automatically when the problem is fixed.

Guiding Light

Midrange and high-end systems, including model 760 and later, are usually repaired by IBM Support personnel.

In the Light Path LED implementation, the system can clearly identify components for replacement by using specific component-level LEDs, and can also guide the servicer directly to the component by signaling (turning on solid) the amber system fault LED, enclosure fault LED, and the component FRU fault LED. The servicer can also use the identify function to blink the FRU-level LED. When this function is activated, a roll-up to the blue enclosure locate

and system locate LEDs will occur. These LEDs will turn on solid and can be used to follow the light path from the system to the enclosure and down to the specific FRU.

Data centers 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 extends to the frame exterior, clearly *guiding* the service representative to the correct rack, system, enclosure, drawer, and component.

Service labels

Service providers use these labels to assist in doing maintenance actions. Service labels are in various formats and positions, and are intended to transmit readily available information to the servicer during the repair process.

Several of these service labels and their purposes are described in the following list:

- Location diagrams are strategically positioned on the system hardware, relating information regarding the placement of hardware components. Location diagrams can include location codes, drawings of physical locations, concurrent maintenance status, or other data that is pertinent to a repair. Location diagrams are especially useful when multiple components are installed, such as DIMMs, sockets, processor cards, fans, adapter cards, LEDs, and power supplies.
- Remove or replace procedure labels contain procedures often found on a cover of the system or in other locations that are accessible to the servicer. These labels provide systematic procedures, including diagrams, detailing how to remove and replace certain serviceable hardware components.
- Numbered arrows are used to indicate the order of operation and serviceability direction of components. Various serviceable parts such as latches, levers, and touch points must be pulled or pushed in a certain direction and order so that the mechanical mechanisms can engage or disengage. Arrows generally improve the ease of serviceability.

The operator panel

The operator panel on a POWER processor-based system is an LCD display (four rows by sixteen elements) that is 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. It includes several buttons, enabling a service support representative (SSR) or client to change various boot-time options and for other limited service functions.

Concurrent maintenance

The IBM POWER7 and POWER7+ 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. Other devices, such as I/O adapters can begin to wear from repeated plugging and unplugging. For these reasons, these devices are specifically designed to be concurrently maintainable when properly configured.

In other cases, a client might be in the process of moving or redesigning a data center or planning a major upgrade. At those times, flexibility is crucial. The IBM POWER7 and POWER7+ processor-based systems are designed for redundant or concurrently maintainable power, fans, physical storage, and I/O towers.

The most recent members of the IBM Power Systems family, based on the POWER7+ processor, continue to support concurrent maintenance of power, cooling, PCI adapters, media devices, I/O drawers, GX adapter, and the operator panel. In addition, they support

concurrent firmware fix pack updates when possible. The determination of whether a firmware fix pack release can be updated concurrently is identified in the readme file that is released with the firmware.

Blind swap cassette

Blind swap PCIe adapters represent significant service and ease-of-use enhancements in I/O subsystem design while maintaining high PCIe adapter density.

Blind swap allows PCIe adapters to be concurrently replaced or installed without having to put the I/O drawer or system into a service position. Since first delivered, minor carrier design adjustments were able to improve an already well-planned service design.

For PCIe adapters on the POWER7+ processor-based servers, blind swap cassettes include the PCIe slot, to avoid the top to bottom movement for inserting the card on the slot that was required on previous designs. The adapter is correctly connected by just sliding the cassette in and actuacting a latch.

Firmware updates

System firmware is delivered as a release level or a service pack. Release levels support the general availability (GA) of new functions or features, and new machine types or models. Upgrading to a higher release level is disruptive to customer operations. IBM intends to introduce no more than two new release levels per year. These release levels will be supported by service packs. Service packs are intended to contain only firmware fixes and not to introduce new function. A *service pack* is an update to an existing release level.

If the system is managed by a management console, you use the management console for firmware updates. By using the management console, you can take advantage of the Concurrent Firmware Maintenance (CFM) option when concurrent service packs are available. CFM is the IBM term used to describe the IBM Power Systems firmware updates that can be partially or wholly concurrent or nondisruptive. With the introduction of CFM, IBM is significantly increasing a client's opportunity to stay on a given release level for longer periods of time. Clients that want maximum stability can defer until there is a compelling reason to upgrade, such as the following reasons:

- A release level is approaching its end-of-service date (that is, it has been available for about a year, and soon, service will not be supported).
- Move a system to a more standardized release level when there are multiple systems in an environment with similar hardware.
- A new release has new functionality that is needed in the environment.
- A scheduled maintenance action will cause a platform reboot, which provides an opportunity to also upgrade to a new firmware release.

The updating and upgrading of system firmware depends on several factors, such as whether the system is stand-alone or managed by a management console, the current firmware installed, and what operating systems are running on the system. These scenarios and the associated installation instructions are comprehensively outlined in the firmware section of Fix Central:

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

You might also want to review the best practice white papers:

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

Repair and verify system

Repair and verify (R&V) is a system that is used to guide a service provider, step-by-step, through the process of repairing a system and verifying that the problem was repaired. The steps are customized in the appropriate sequence for the particular repair for the specific system being repaired. The following scenarios are covered by repair and verify:

- Replacing a defective field-replaceable unit (FRU) or a customer-replaceable unit (CRU)
- 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

Repair and verify procedures can be used by service representative providers who are familiar with the task and those who are not. Education-on-demand 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.

If a server is managed by a management console, then many of the repair and verify procedures are done from the management console. If the FRU to be replaced is a PCI adapter or an internal storage device, the service action is always performed from the operating system of the partition owning that resource.

Clients can subscribe through the subscription services to obtain the notifications about the latest updates available for service-related documentation. The latest version of the documentation is accessible through the Internet.

4.4 Manageability

Several functions and tools help manageability so you can efficiently and effectively manage your system.

4.4.1 Service user interfaces

The service interface allows support personnel or the client to communicate with the service support applications in a server by using a console, interface, or terminal. Delivering a clear, concise view of available service applications, the service interface allows the support team to manage system resources and service information in an efficient and effective way. Applications that are available through the service interface are carefully configured and placed to give service providers access to important service functions.

Various service interfaces are used, depending on the state of the system and its operating environment. The primary service interfaces are the following items:

- Light Path and Guiding Light

  See “Light Path” on page 158 and “Guiding Light” on page 158.
- Service processor, Advanced System Management Interface (ASMI)
- Operator panel
- Operating system service menu
- Service Focal Point on the Hardware Management Console
- Service Focal Point Lite on Integrated Virtualization Manager

Service processor

The service processor is a controller that is 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 that is connected to the POWER processor. The service processor is always working, regardless of the 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 that the connection to the management console 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 by using the ASMI, and enables complete system and partition management from the management console.

With two CEC enclosures and more, there are two redundant FSPs, one in each of the first CECs. While one is active, the second one is in standby mode. In case of a failure, there is automatic takeover.

Analyze system that does not boot: The service processor enables a system that does not boot to be analyzed. The error log analysis can be done from either the ASMI or the management console.

The service processor uses two Ethernet ports that run at 100 Mbps speed. Consider the following information:

- Both Ethernet ports are visible only 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 support only auto-negotiation. Customer-selectable media speed and duplex settings are not available.
- Both Ethernet ports have a default IP address, as follows:
  - Service processor eth0 (HMC1 port) is configured as 169.254.2.147.
  - Service processor eth1 (HMC2 port) is configured as 169.254.3.147.

The following functions are available through service processor:

- Call home
- Advanced System Management Interface (ASMI)
- Error Information (error code, part number, location codes) menu
- View of guarded components
- Limited repair procedures
- Generate dump
- LED Management menu
- Remote view of ASMI menus
- Firmware update through USB key

Advanced System Management Interface

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. Various repair procedures require connection to the ASMI.

The ASMI is accessible through the management console. It is also accessible by 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. ASMI can also be accessed from an ASCII terminal, but this is only available while the system is in the platform powered-off mode.

Use the ASMI to change the service processor IP addresses or to apply certain 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. To access ASMI, use one of the following methods:

- Use a management console.

  If configured to do so, the management console connects directly to the ASMI for a selected system from this task.

  To connect to the Advanced System Management interface from a management console, use the following steps:

  a. Open **Systems Management** from the navigation pane.

  b. From the work panel, select one or more managed systems to work with.

  c. From the System Management tasks list, select **Operations Advanced System Management** (ASM).

- Use a web browser.

  At the time of writing, supported web browsers are Microsoft Internet Explorer (Version 10.0.9200.16439), Mozilla Firefox (Version 17.0.2), and Opera (Version 9.24). Later versions of these browsers might work but are not officially supported. The JavaScript language and cookies must be enabled.

  The web interface is available during all phases of system operation, including the initial program load (IPL) and run time. However, several 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 by using the following address:

  `https://<ip_address_of_service_processor>`

  Note: To make the connection through Internet Explorer, click **Tools Internet Options**. Clear the **Use TLS 1.0** check box, and click **OK**.

- Use an ASCII terminal.

  The ASMI on an ASCII terminal supports a subset of the functions that are provided by the web interface and is available only when the system is in the platform powered-off mode. The ASMI on an ASCII console is not available during several phases of system operation, such as the IPL and run time.

The operator panel

The service processor provides an interface to the operator panel, which is used to display system status and diagnostic information.

The operator panel can be accessed in two ways:

- By using the normal operational front view.
- By pulling it out to access the switches and viewing the LCD display.

Figure 4-5 shows that the operator panel on a Power 710 and Power 730 is pulled out

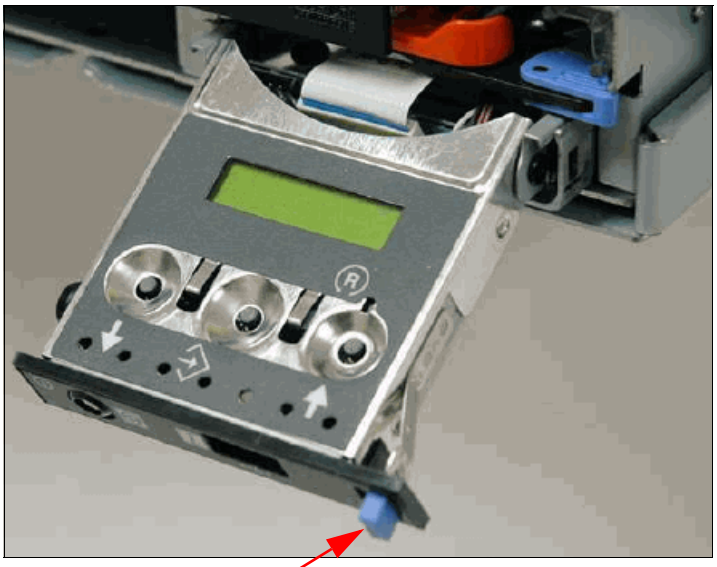


Figure 4-5 Operator panel is pulled out from the chassis

Several of the operator panel features include the following items:

- A 2 x 16 character LCD display
- Reset, enter, power On/Off, increment, and decrement buttons
- Amber System Information/Attention, green Power LED
- Blue Enclosure Identify LED on the Power 710 and Power 730
- Altitude sensor
- USB Port
- Speaker/Beeper

The following functions are available through the operator panel:

- Error Information
- Generate dump
- View machine type, model, and serial number
- Limited set of repair functions

Operating system service menu

The system diagnostics consist of IBM i service tools, stand-alone diagnostics that are loaded from the DVD drive, and online diagnostics (available in AIX).

Online diagnostics, when installed, are a part of the AIX or IBM i 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. IBM i has a service tools problem log, IBM i history log (QHST), and IBM i problem log.

The modes are as follows:

- Service mode

  This mode requires a service mode boot of the system and enables the checking of system devices and features. Service mode provides the most complete self-check of the system resources. All system resources, except the SCSI adapter and the disk drives used for paging, can be tested.

- Concurrent mode

  This mode enables the normal system functions to continue while selected resources are being checked. Because the system is running in normal operation, certain devices might require additional actions by the user or diagnostic application before testing can be done.

- Maintenance mode

  This 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 that 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.

Alternate method: When you order a Power System, a DVD-ROM or DVD-RAM might be 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 IBM i operating system and associated machine code 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 keeps all applications running, by 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 other options.

Depending on the operating system, the following service-level functions are what you typically see when you use the operating system service menus:

- 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
- Call Home/Customer Notification
- Error information menu
- LED management menu
- Concurrent/Non-concurrent maintenance (within scope of the OS)
- Managing firmware levels
  - Server
  - Adapter
- Remote support (access varies by OS)

Service Focal Point on the Hardware Management Console

Service strategies become more complicated in a partitioned environment. The Manage Serviceable Events task in the management console can help to streamline this process.

Each logical partition reports errors that it detects and forwards the event to the Service Focal Point (SFP) application that is running on the management console, 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 management console, 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.

4.4.2 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 and associated I/O adapters.

The system firmware consists of service processor microcode, Open Firmware microcode, SPCN microcode, and the POWER Hypervisor.

The firmware and microcode can be downloaded and installed either from an HMC, from a running partition, or from USB port number 1 on the rear of a Power 710 and Power 730, if that system is not managed by an HMC.

Power Systems has a permanent firmware boot side (A side) and a temporary firmware boot side (B side). New levels of firmware must be installed first on the temporary side to test the update's compatibility with existing applications. When the new level of firmware is approved, it can be copied to the permanent side.

For access to the initial web pages that address this capability, see the Support for IBM Systems web page:

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

For Power Systems, select the **Power** link. Figure 4-6 shows an example.

Figure 4-6 Support for Power servers web page

Although the content under the Popular links section can change, click the **Firmware and HMC updates** link to go to the resources for keeping your system's firmware current.

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

Each IBM Power Systems server has the following levels of server firmware and power subsystem firmware:

- Installed level

  This level of server firmware or power subsystem firmware is installed and will be installed into memory after the managed system is powered off and then powered on. It is installed on the temporary side of system firmware.

- Activated level

  This level of server firmware or power subsystem firmware is active and running in memory.

- Accepted level

  This level 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 permanent side of system firmware.

IBM provides the Concurrent Firmware Maintenance (CFM) function on selected Power Systems. This function supports applying nondisruptive system firmware service packs to the system concurrently (without requiring a reboot operation 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. Deferred fixes, if any, will be identified in the Firmware Update Descriptions table of the firmware document. For deferred fixes within a service pack, only the fixes in the service pack that cannot be concurrently activated are deferred. Table 4-1 shows the file-naming convention for system firmware.

Table 4-1 Firmware naming convention

PPNNSSS_FFF_DDD			
PP	Package identifier	01	-
		02	-
NN	Platform and class	AL	Low end
		AM	Mid range
		AS	Blade server
		AH	High end
		AP	Bulk power for IH
		AB	Bulk power for high end
SSS	Release indicator		
FFF	Current fix pack		
DDD	Last disruptive fix pack		

The following example uses the convention:

```
01AL770_032 = POWER7+ Entry Systems Firmware for 8231-E1D and 8231-E2D
```

Support for PowerLinux: The minimum firmware level 01AL770_032 will also be used on the PowerLinux 7R1 (8246-L1D and 8246-L1T) and PowerLinux 7R2 (8246-L2D and 8246-L2T) systems.

An installation is disruptive if the following statements are true:

- The release levels (SSS) of currently installed and new firmware differ.
- 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.

4.4.3 Concurrent firmware update improvements with POWER7+

Since POWER6, firmware service packs are generally concurrently applied and take effect immediately. Occasionally, a service pack is shipped where most of the features can be concurrently applied; but because changes to some server functions (for example, changing initialization values for chip controls) cannot occur during operation, a patch in this area required a system reboot for activation.

With the Power-On Reset Engine (PORE), the firmware can now dynamically power off processor components, make changes to the registers and re-initialize while the system is running, without discernible impact to any applications running on a processor. This potentially allows concurrent firmware changes in POWER7+, which in earlier designs, required a reboot to take effect.

Activating some new firmware functions requires installation of a firmware release level. This process is disruptive to server operations and requires a scheduled outage and full server reboot.

4.4.4 Electronic Services and Electronic Service Agent

IBM transformed its delivery of hardware and software support services to help you achieve higher system availability. Electronic Services is a web-enabled solution that offers an exclusive, no-additional-charge enhancement to the service and support that is available for IBM servers. These services provide the opportunity for greater system availability with faster problem resolution and preemptive monitoring. The Electronic Services solution consists of two separate, but complementary, elements:

- Electronic Services news page

  The Electronic Services news page is a single Internet entry point that replaces the multiple entry points, which are traditionally used to access IBM Internet services and support. With the news page, you can gain easier access to IBM resources for assistance in resolving technical problems.

- Electronic Service Agent

  The Electronic Service Agent is software that resides on your server. It monitors events and transmits system inventory information to IBM on a periodic, client-defined timetable. The Electronic Service Agent automatically reports hardware problems to IBM.

Early knowledge about potential problems enables IBM to deliver proactive service that can result in higher system availability and performance. In addition, information that is collected

through the Service Agent is made available to IBM service support representatives when they help answer your questions or diagnose problems. Installation and use of IBM Electronic Service Agent for problem reporting enables IBM to provide better support and service for your IBM server.

To learn how Electronic Services can work for you, visit the following site; an IBM ID is required:

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

Benefits are as follows:

- Increased uptime

  The Electronic Service Agent tool is designed to enhance the warranty or maintenance agreement by providing faster hardware error reporting and uploading system information to IBM Support. This way can translate to less wasted time monitoring the symptoms, diagnosing the error, and manually calling IBM Support to open a problem record.

  Its 24x7 monitoring and reporting mean no more dependence on human intervention or off-hours customer personnel when errors are encountered in the middle of the night.

- Security

  The Electronic Service Agent tool is designed to be secure in monitoring, reporting, and storing the data at IBM. The Electronic Service Agent tool securely transmits either with the Internet (HTTPS or VPN) or modem, and can be configured to communicate securely through gateways to provide customers a single point of exit from their site.

  Communication is one way. Activating Electronic Service Agent does not enable IBM to call into a customer's system. System inventory information is stored in a secure database, which is protected behind IBM firewalls. It is viewable only by the customer and IBM. The customer's business applications or business data is never transmitted to IBM.

- More accurate reporting

  Because system information and error logs are automatically uploaded to the IBM Support center in conjunction with the service request, customers are not required to find and send system information, decreasing the risk of misreported or misdiagnosed errors.

  When inside IBM, problem error data is run through a data knowledge management system and knowledge articles are appended to the problem record.

- Customized support

  By using the IBM ID that you enter during activation, you can view system and support information by selecting **My Systems** at the Electronic Support website:

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

  My Systems provides valuable reports of installed hardware and software, using information collected from the systems by Electronic Service Agent. Reports are available for any system associated with the customers IBM ID. Premium Search combines the function of search and the value of Electronic Service Agent information, providing advanced search of the technical support knowledge base. Using Premium Search and the Electronic Service Agent information that was collected from your system, your clients are able to see search results that apply specifically to their systems.

  For more information about how to use the power of IBM Electronic Services, contact your IBM Systems Services Representative, or visit the following website:

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

4.5 POWER7+ RAS features

This section lists POWER7+ RAS features in this release:

- Power-On Reset Engine (PORE)

  Enables a processor to be re-initialized while the system remains running. This feature will allow for the concurrent firmware updates situation, in which a processor initialization register value needs to be changed. Concurrent firmware updates might be more prevalent.

- L3 Cache dynamic column repair

  This self-healing capability completes cache-line delete and uses the PORE feature to potentially avoid some repair actions or outages that are related to L3 cache.

- Accelerator RAS

  New accelerators are designed with RAS features to avoid system outages in the vast majority of faults that can be detected by the accelerators.

- Fabric Bus Dynamic Lane Repair

  POWER7+ has spare bit lanes that can dynamically be repaired (using PORE). This feature avoids any repair action or outage related to a single bit failure for the fabric bus.

4.6 Power-On Reset Engine

The POWER7+ chip includes a Power-On Reset Engine (PORE), a programmable hardware sequencer responsible for restoring the state of a powered down processor core and L2 cache (deep sleep mode), or chiplet (winkle mode). When a processor core wakes up from sleep or winkle, the PORE fetches code created by the POWER Hypervisor from a special location in memory containing the instructions and data necessary to restore the processor core to a functional state. This memory image includes all the necessary boot and runtime configuration data that were applied to this processor core since power-on, including circuit calibration and cache repair registers that are unique to each processor core. Effectively the PORE performs a mini initial program load (IPL) of the processor core or chiplet, completing the sequence of operations necessary to restart instruction execution, such as removing electrical and logical fences and reinitializing the Digital PLL clock source.

Because of its special ability to perform clocks-off and clocks-on sequencing of the hardware, the PORE can also be used for RAS purposes:

- The service processor can use the PORE to concurrently apply an initialization update to a processor core/chiplet by loading new initialization values into memory and then forcing it to go in and out of winkle mode. This step happens, all without causing disruption to the workloads or operating system (all occurring in a few milliseconds).
- In the same fashion, PORE can initiate an L3 cache dynamic “bit-line” repair operation if the POWER Hypervisor detects too many recoverable errors in the cache.
- The PORE can be used to dynamically repair node-to-node fabric bit lanes in a POWER7+ processor-based server by quickly suspending chip-chip traffic during run time, reconfiguring the interface to use a spare bit lane, then resuming traffic, all without causing disruption to the operation of the server.

4.7 Operating system support for RAS features

Table 4-2 gives an overview of features for continuous availability that are supported by the various operating systems running on power systems. In the table, the word “Most” means most functions.

Table 4-2 Operating system support for RAS features

RAS feature	AIX 5.3	AIX 6.1	AIX 7.1	IBM i	RHEL 5.7	RHEL 6.3	SLES11 SP2
System deallocation of failing components							
Processor Fabric Bus Protection	X	X	X	X	X	X	X
Dynamic Processor Deallocation	X	X	X	X	X	X	X
Dynamic Processor Sparing	X	X	X	X	X	X	X
Processor Instruction Retry	X	X	X	X	X	X	X
Alternate Processor Recovery	X	X	X	X	X	X	X
Partition Contained Checkstop	X	X	X	X	X	X	X
Persistent processor deallocation	X	X	X	X	X	X	X
GX++ bus persistent deallocation	X	X	X	X	-	-	X
Optional ECC I/O hub with freeze behavior	X	X	X	X	X	X	X
PCI bus extended error detection	X	X	X	X	X	X	X
PCI bus extended error recovery	X	X	X	X	Most	Most	Most
PCI-PCI bridge Enhanced Error Handling	X	X	X	X	-	-	-
Redundant RIO or 12x Channel link[a]	X	X	X	X	X	X	X
PCI card hot-swap	X	X	X	X	X	X	X
Dynamic SP failover at run time[b]	X	X	X	X	X	X	X
Memory sparing with CoD at IPL time	X	X	X	X	X	X	X
Clock failover run time or IPL[b]	X	X	X	X	X	X	X
Memory availability							
ECC memory, L2, L3 cache	X	X	X	X	X	X	X
CRC plus retry on memory data bus	X	X	X	X	X	X	X
Data Bus	X	X	X	X	X	X	X
Dynamic memory channel repair	X	X	X	X	X	X	X
Processor memory controller memory scrubbing	X	X	X	X	X	X	X
Memory page deallocation	X	X	X	X	X	X	X
Chipkill memory	X	X	X	X	X	X	X
L1 instruction and data array protection	X	X	X	X	X	X	X
L2/L3 ECC and cache line delete	X	X	X	X	X	X	X

RAS feature	AIX 5.3	AIX 6.1	AIX 7.1	IBM i	RHEL 5.7	RHEL 6.3	SLES11 SP2
Special uncorrectable error handling	X	X	X	X	X	X	X
Active Memory Mirroring for Hypervisor[b]	X	X	X	X	X	X	X
Fault detection and isolation							
Platform FFDC diagnostics	X	X	X	X	X	X	X
Run-time diagnostics	X	X	X	X	Most	Most	Most
Storage Protection Keys	-	X	X	X	-	-	-
Dynamic Trace	X	X	X	X	-	-	X
Operating System FFDC	-	X	X	X	-	-	-
Error log analysis	X	X	X	X	X	X	X
Freeze mode of I/O Hub	X	X	X	X	-	-	-
Service processor support for:							
► Built-in self-tests (BIST) for logic and arrays	X	X	X	X	X	X	X
► Wire tests	X	X	X	X	X	X	X
► Component initialization	X	X	X	X	X	X	X
Serviceability							
Boot-time progress indicators	X	X	X	X	Most	Most	Most
Electronic Service Agent Call Home from management console	X	X	X	X	X	X	X
Firmware error codes	X	X	X	X	X	X	X
Operating system error codes	X	X	X	X	Most	Most	Most
Inventory collection	X	X	X	X	X	X	X
Environmental and power warnings	X	X	X	X	X	X	X
Hot-plug fans, power supplies	X	X	X	X	X	X	X
Extended error data collection	X	X	X	X	X	X	X
I/O drawer redundant connections[a]	X	X	X	X	X	X	X
I/O drawer hot add and concurrent repair[a]	X	X	X	X	X	X	X
HOT GX adapter add and repair[b]	X	X	X	X	X	X	X
Concurrent add of powered I/O rack[a]	X	X	X	X	X	X	X
SP mutual surveillance with POWER Hypervisor	X	X	X	X	X	X	X
Dynamic firmware update with management console	X	X	X	X	X	X	X
PORE: Core Inizialization without reboot	X	X	X	X	X	X	X
Service processor support for BIST	X	X	X	X	X	X	X
Electronic Service Agent Call Home Application	X	X	X	X	-	-	-

RAS feature	AIX 5.3	AIX 6.1	AIX 7.1	IBM i	RHEL 5.7	RHEL 6.3	SLES11 SP2
Guiding light LEDs	X	X	X	X	X	X	X
System dump for memory, POWER Hypervisor, SP	X	X	X	X	X	X	X
Information center / Systems Support Site service publications	X	X	X	X	X	X	X
System Support Site education	X	X	X	X	X	X	X
Operating system error reporting to management console SFP	X	X	X	X	X	X	X
RMC secure error transmission subsystem	X	X	X	X	X	X	X
Health check scheduled operations with management console	X	X	X	X	X	X	X
Operator panel (real or virtual)	X	X	X	X	X	X	X
Concurrent operator panel maintenance[b]	X	X	X	X	X	X	X
Redundant management consoles	X	X	X	X	X	X	X
Automated server recovery/restart	X	X	X	X	X	X	X
PowerVM Live Partition Mobility	X	X	X	X	X	X	X
Live Application Mobility	-	X	X	-	-	-	-
Repair and Verify Guided Maintenance	X	X	X	X	Most	Most	Most
Concurrent kernel update	-	X	X	X	X	X	X
Concurrent Hot Add/Repair Maintenance	X	X	X	X	X	X	X
Power and cooling							
Redundant, hot swap fans and blower for CEC	X	X	X	X	X	X	X
Redundant, hot swap power for CEC	X	X	X	X	X	X	X
TPMD for system power and thermal management	X	X	X	X	X	X	X
CEC power/thermal sensor (CPU and memory)	X	X	X	X	X	X	X
Redundant power for I/O drawers[a]	X	X	X	X	X	X	X

a. Not available on Power 710 and Power 730.

b. Need mid-tier and large-tier POWER7 systems or later, including Power 770, 780, and 795.

Related publications

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

IBM Redbooks

The following IBM Redbooks publications provide additional information about the topic in this document. Note that some publications referenced in this list might be available in softcopy only.

- *An Introduction to Fibre Channel over Ethernet, and Fibre Channel over Convergence Enhanced Ethernet*, REDP-4493
- *IBM BladeCenter PS700, PS701, and PS702 Technical Overview and Introduction*, REDP-4655
- *IBM BladeCenter PS703 and PS704 Technical Overview and Introduction*, REDP-4744
- *IBM Power 720 and 740 (8202-E4D, 8205-E6D) Technical Overview and Introduction*, REDP-4984
- *IBM Power 750 and 760 (8408-E8D, 9109-RMD) Technical Overview and Introduction*, REDP-4985
- *IBM Power 750 and 755 (8233-E8B, 8236-E8C) Technical Overview and Introduction*, REDP-4638
- *IBM Power 770 and 780 (9117-MMD, 9179-MHD) Technical Overview and Introduction*, REDP-4798
- *IBM Power 795 (9119-FHB) Technical Overview and Introduction*, REDP-4640
- *IBM Power Systems HMC Implementation and Usage Guide*, SG24-7491
- *IBM Power Systems: SDMC to HMC Migration Guide (RAID1)*, REDP-4872
- *IBM PowerVM Virtualization Introduction and Configuration*, SG24-7940
- *IBM PowerVM Virtualization Managing and Monitoring*, SG24-7590
- *IBM PowerVM Best Practices*, SG24-8062
- *IBM PowerVM Live Partition Mobility*, SG24-7460
- *IBM Systems Director 6.3 Best Practices: Installation & Configuration*, REDP-4932
- *PowerVM Migration from Physical to Virtual Storage*, SG24-7825
- *PowerVM and SAN Copy Services*, REDP-4610

You can search for, view, download or order these documents and other Redbooks, Redpapers, Web Docs, draft and additional materials, at the following website:

ibm.com/redbooks

© Copyright IBM Corp. 2013. All rights reserved.

Other publications

These publications are also relevant as further information sources:

- IBM Power Facts and Features: IBM Power Systems, IBM PureFlex and Power Blades
  http://www.ibm.com/systems/power/hardware/reports/factsfeatures.html
- Specific storage devices supported for Virtual I/O Server
  http://www14.software.ibm.com/webapp/set2/sas/f/vios/documentation/datasheet.html
- IBM Power 710 server data sheet
  http://public.dhe.ibm.com/common/ssi/ecm/en/pod03048usen/POD03048USEN.PDF
- IBM Power 720 server data sheet
  http://public.dhe.ibm.com/common/ssi/ecm/en/pod03049usen/POD03049USEN.PDF
- IBM Power 730 server data sheet
  http://public.dhe.ibm.com/common/ssi/ecm/en/pod03050usen/POD03050USEN.PDF
- IBM Power 740 server data sheet
  http://public.dhe.ibm.com/common/ssi/ecm/en/pod03051usen/POD03051USEN.PDF
- IBM Power 750 server data sheet
  http://public.dhe.ibm.com/common/ssi/ecm/en/pod03034usen/POD03034USEN.PDF
- IBM Power 755 server data sheet
  http://public.dhe.ibm.com/common/ssi/ecm/en/pod03035usen/POD03035USEN.PDF
- IBM Power 760 server data sheet
  http://public.dhe.ibm.com/common/ssi/ecm/en/pod03080usen/POD03080USEN.PDF
- IBM Power 770 server data sheet
  http://public.dhe.ibm.com/common/ssi/ecm/en/pod03031usen/POD03031USEN.PDF
- IBM Power 780 server data sheet
  http://public.dhe.ibm.com/common/ssi/ecm/en/pod03032usen/POD03032USEN.PDF
- IBM Power 795 server data sheet
  http://public.dhe.ibm.com/common/ssi/ecm/en/pod03053usen/POD03053USEN.PDF
- Active Memory Expansion: Overview and Usage Guide
  http://public.dhe.ibm.com/common/ssi/ecm/en/pow03037usen/POW03037USEN.PDF
- POWER7 System RAS Key Aspects of Power Systems Reliability, Availability, and Serviceability
  http://public.dhe.ibm.com/common/ssi/ecm/en/pow03056usen/POW03056USEN.PDF

Online resources

These websites are also relevant as further information sources:

- IBM Power Systems Hardware Information Center

  http://pic.dhe.ibm.com/infocenter/powersys/v3r1m5/index.jsp
- IBM System Planning Tool website

  http://www.ibm.com/systems/support/tools/systemplanningtool/
- IBM Fix Central website

  http://www.ibm.com/support/fixcentral/
- Power Systems Capacity on Demand website

  http://www.ibm.com/systems/power/hardware/cod/
- Support for IBM Systems website

  http://www.ibm.com/support/entry/portal/Overview?brandind=Hardware~Systems~Power
- IBM Power Systems website

  http://www.ibm.com/systems/power/
- IBM Storage website

  http://www.ibm.com/systems/storage/
- IBM Systems Energy Estimator

  http://www-912.ibm.com/see/EnergyEstimator/
- Migration combinations of processor compatibility modes for active Partition Mobility

  http://publib.boulder.ibm.com/infocenter/powersys/v3r1m5/topic/p7hc3/iphc3pcmcombosact.htm
- Advance Toolchain:

  http://ibm.co/1O6nMYI

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services

IBM Power 710 and 730 Technical Overview and Introduction

Features 8231-E1D, 8231-E2D, 8246 PowerLinux servers based on POWER7+ processor technology

Describes the support of 20 partitions per core

Explains 2U rack-mount design for leading performance

This IBM Redpaper publication is a comprehensive guide covering the IBM Power 710 (8231-E1D) and Power 730 (8231-E2D) servers that support IBM AIX, IBM i, and Linux operating systems. This paper also describes the IBM PowerLinux 7R1 (8246-L1D and 8246-L1T) and the PowerLinux 7R2 (8246-L2D and 8246-L2T) servers that support the Linux operating system. The goal of this paper is to introduce the innovative Power 710, Power 730, PowerLinux 7R1, and PowerLinux offerings and their major functions:

- IBM POWER7+ processor is available at frequencies of 3.6 GHz, 4.2 GHz, and 4.3 GHz.
- Larger IBM POWER7+ Level 3 cache provides greater bandwidth, capacity, and reliability.
- Integrated SAS/SATA controller for HDD, SSD, tape, and DVD supports built-in hardware RAID 0, 1, and 10.
- New IBM PowerVM V2.2.2 features, such as 20 LPARs per core.
- Improved IBM Active Memory Expansion technology provides more usable memory than is physically installed in the system.

Professionals who want to acquire a better understanding of IBM Power Systems products can benefit from reading this paper.

This paper expands the current set of IBM Power Systems documentation by providing a desktop reference that offers a detailed technical description of the Power 710 and Power 730 systems.

This paper does not replace the latest marketing materials and configuration 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-4983-00

INTERNATIONAL TECHNICAL SUPPORT ORGANIZATION

BUILDING TECHNICAL INFORMATION BASED ON PRACTICAL EXPERIENCE

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

For more information:
ibm.com/redbooks