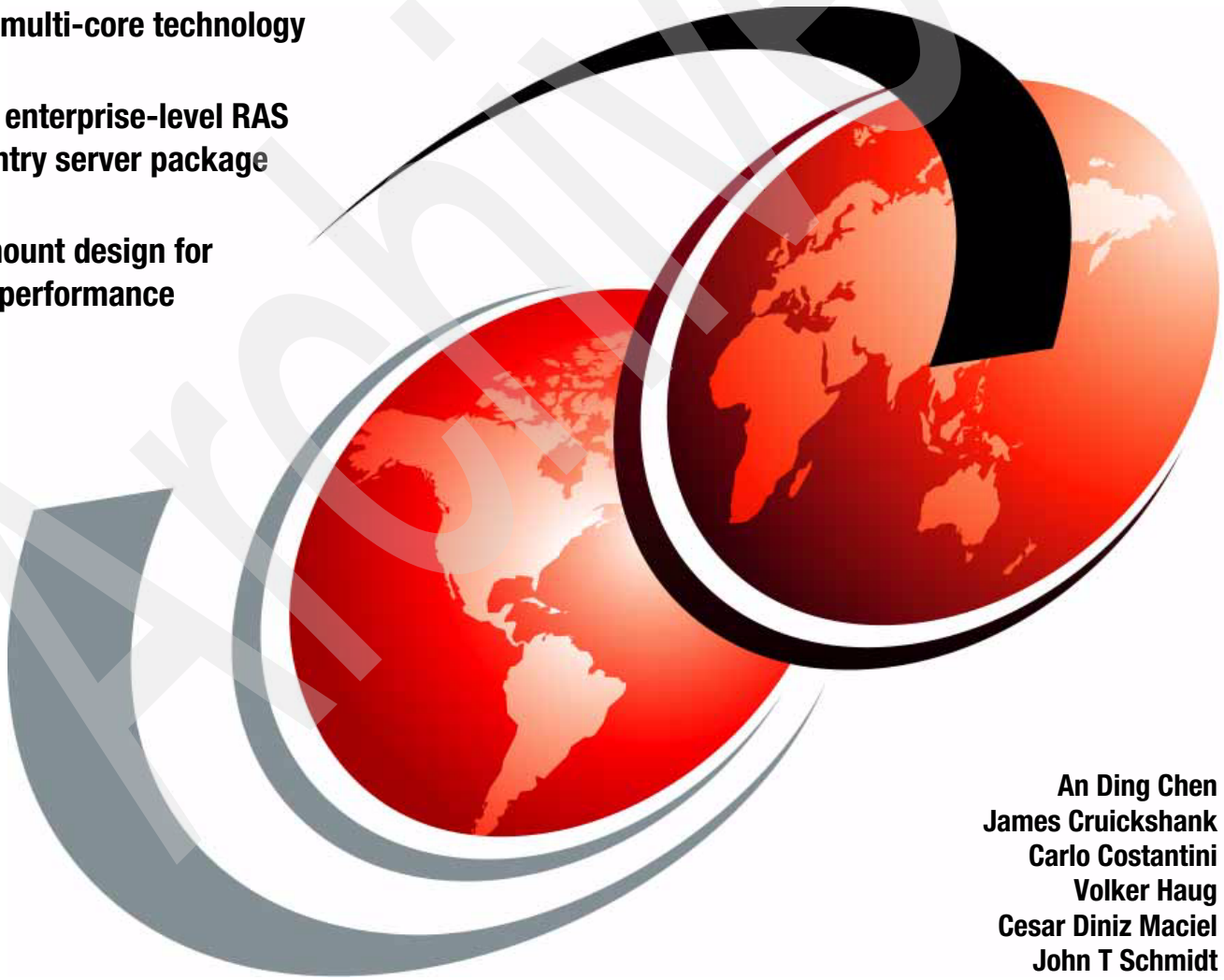
IBM

IBM Power 710 and 730

Technical Overview and Introduction

- Features the POWER7 processor providing advanced multi-core technology
- PowerVM, enterprise-level RAS all in an entry server package
- 2U rack-mount design for midrange performance



An Ding Chen
James Cruickshank
Carlo Costantini
Volker Haug
Cesar Diniz Maciel
John T Schmidt

Redpaper

ibm.com/redbooks

International Technical Support Organization

IBM Power 710 and 730: Technical Overview and Introduction

October 2010

REDP-4636-00

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

First Edition (October 2010)

This edition applies to the IBM Power 710 and Power 730 (8231-E2B) Power Systems servers.

This document created or updated on March 21, 2012.

Contents

Notices

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

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

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

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

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

Any references in this information to non-IBM 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.

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

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

COPYRIGHT LICENSE:

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

Trademarks

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

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

Active Memory™
AIX 5L™
AIX®
DB2®
DS8000®
Electronic Service Agent™
EnergyScale™
Focal Point™
IBM Systems Director Active Energy Manager™
IBM®
Micro-Partitioning™
Power Architecture®
POWER Hypervisor™
Power Systems™
Power Systems Software™
POWER4™
POWER4+™
POWER5™
POWER5+™
POWER6+™
POWER6®
POWER7™
PowerHA™
PowerPC®
PowerVM™
POWER®
pSeries®
pureScale™
Rational Team Concert™
Rational®
Redbooks®
Redpaper™
Redbooks (logo) ®
RS/6000®
System p5®
System p®
System Storage®
System z®
Tivoli®
Workload Partitions Manager™
XIV®

The following terms are trademarks of other companies:

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

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

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

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

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

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

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

Preface

This IBM® Redpaper™ publication is a comprehensive guide covering the IBM Power 710 and Power 730 servers supporting IBM AIX®, IBM i, and Linux® operating systems. The goal of this paper is to introduce the innovative Power 710 and Power 730 offerings and their major functions, including these:

- The IBM POWER7™ processor available at frequencies of 3.0 GHz, 3.55 GHz, and 3.7 GHz
- The specialized POWER7 Level 3 cache that provides greater bandwidth, capacity, and reliability
- The 1 Gb or 10 Gb Integrated Virtual Ethernet adapter, included with each server configuration, providing native hardware virtualization
- The latest IBM PowerVM™ virtualization, including PowerVM Live Partition Mobility and PowerVM IBM Active Memory™ Sharing
- 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

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

The team who wrote this paper

This paper was produced by a team of specialists from around the world working at the International Technical Support Organization, Austin Center, Texas. The project that produced this publication was managed by Scott Vetter.

An Ding Chen is a Power systems Product Engineer in Shanghai China, who provides level 3 hardware and firmware support in all Asia Pacific countries and Japan. He has ten years experience on AIX/UNIX®, IBM RS/6000®, IBM pSeries®, and Power Systems products. Starting from his university, he has passed the CATE certification on pSeries systems and IBM AIX 5L™. He joined IBM in 2006.

James Cruickshank works on the Power Systems Field Technical Sales Support team for IBM in the UK. He holds an honors degree in Mathematics from the University of Leeds. James has nine years of experience working on RS/6000, 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.

Carlo Costantini is a Certified IT Specialist for IBM and has over 32 years of experience with IBM and IBM Business Partners. He currently works in Italy Power Systems Platforms as Presales Field Technical Sales Support for IBM Sales Representatives and IBM Business Partners. Carlo has broad marketing experience and his current major areas of focus are competition, sales, and technical sales support. He is a certified specialist for Power Systems servers. He holds a Masters degree in Electronic Engineering from Rome University.

Volker Haug is a certified Consulting IT Specialist within IBM Systems and Technology Group, based in Ehningen, Germany. He holds a Bachelor's degree in Business Management from the university of Applied Studies in Stuttgart. His career has included more than 23 years working in the IBM PLM and Power Systems divisions as a RISC and AIX Systems Engineer. Volker is an expert in Power Systems hardware, AIX, and PowerVM virtualization. He is a member of the EMEA Power Champions team and also a member of the German Technical Expert Council, a affiliate of the IBM Academy of Technology. He has written various books and white papers about AIX, workstations, servers, and PowerVM virtualization.

Cesar Diniz Maciel is a Senior Consulting IT Specialist with IBM in the United States. He joined IBM in 1996 as Presales Technical Support for the RS/6000 family of UNIX servers in Brazil, and came to IBM United States in 2005. He is part of the Global Techline team, working on presales consulting for Latin America. He holds a degree in Electrical Engineering from Universidade Federal de Minas Gerais (UFMG) in Brazil. His areas of expertise include Power Systems, AIX, and IBM POWER® Virtualization. He has written extensively on Power Systems and related products. This is his seventh ITSO residency.

John T Schmidt is an Accredited IT Specialist for IBM and has ten years 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 2010, he completed an assignment with the IBM Corporate Service Corps in Hyderabad, India. He is currently working in the United States as a presales Field Technical Sales Specialist for Power Systems in St. Louis, MO.

The project that produced this publication was managed by:

Scott Vetter IBM certified project manager and PMP.

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

Salim Agha, Ray W. Anderson, Richard W. Bishop, Joe Cahill, Scott A. Carroll, Sertac Cakici, Hsien-I Chang, Binh Chu, Michael Floyd, George Gaylord, Raymond J. Harrington, Daniel Henderson, John Hilburn,Tenley Jackson, Bob Kovacs, Walter M. Lipp, Mary E. Magnuson, Bill McWaters, Duc T. Nguyen, Thoi Nguyen, Guy R. Paradise, David Pirnik, Todd J. Rosedahl, Pat O'Rourke, Amartey S. Pearson, Jeff Scheel, Joseph C. Schiavone, Timothy Stack, Vasu Vallabhaneni, Jesus G. Villarreal
IBM U.S.A.

Sabine Jordan, Stephen Lutz
IBM Germany

Tamikia Barrow, Brooke C Gardner
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 publications

- 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 publications 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 and IBM Power 730 servers (8231-E2B) utilize the latest POWER7 processor technology designed to deliver unprecedented performance, scalability, reliability, and manageability for demanding commercial workloads.

The Power 710 server is a high-performance, energy efficient, reliable and secure infrastructure and application server in a dense form factor. As a high-performance infrastructure or application server, the Power 710 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 resulting in one of the most cost-efficient solutions for UNIX, 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 making use of the IBM industrial strength PowerVM technology to fully utilize the server's capability.

1.1 Systems

Figure 1-1 shows the Power 710 (top) and Power 730 (bottom).

Figure 1-1 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 3.0 GHz, 6-core 3.7 GHz, and 8-core 3.55 GHz configurations. The POWER7 processor chips in this server are 64-bit, 4-core, 6-core and 8-core modules with 4 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 allows for a maximum system memory of 64 GB.

The Power 710 server offers three storage backplane options. The first supports three Small Form Factor (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/SSDs and a SATA DVD. The third supports six SFF SAS HDDs/SSDs, a SATA DVD, Dual Write Cache RAID, and an external SAS port. HDDs/SSDs are hot-swap and front accessible.

The Power 710 comes with four PCI Express (PCIe) low profile (LP) slots for installing adapters in the system. Also available in the Power 710 system units is a choice of quad gigabit or dual 10 Gb Integrated Virtual Ethernet (IVE) adapters. These native ports can be selected at the time of initial order. Virtualization of these integrated Ethernet adapters is supported.

1.1.2 The Power 730 server

The IBM Power 730 server is a 2U rack-mount server with two processor sockets offering 8-core 3.0 and 3.7 GHz, 12-core 3.7 GHz, and 16-core 3.55 GHz configurations. The POWER7 processor chips in this server are 64-bit, 4-core, 6-core, and 8-core modules with 4 MB of L3 cache per core and 256 KB of L2 cache per core.

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 supporting four DIMM slots, can be installed, allowing maximum system memory of 128 GB.

The Power 730 server offers three storage backplane options. The first supports three SFF SAS hard disk drives (HDDs) or solid state drives (SSDs), an SATA DVD, and a half-high tape drive. The second supports six SFF SAS HDDs/SSDs and an SATA DVD. The third supports six SFF SAS HDDs/SSDs, an SATA DVD, Dual Write Cache RAID, and an external SAS port. HDDs/SSDs are hot-swap and front accessible.

The Power 730 comes with four PCI Express (PCIe) low profile (LP) slots for installing adapters in the system. Also available in the Power 730 system units is a choice of quad gigabit or dual 10 Gb Integrated Virtual Ethernet (IVE) adapters. These native ports can be selected at the time of initial order. Virtualization of these integrated Ethernet adapters is supported.

1.2 Operating environment

The operating environment specifications for the servers can be seen in Table 1-1.

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 to 35 degrees C (41 to 95 degrees F) Preferable: 18 to 27 degrees C (64 to 80 degrees F)		5 to 45 degrees C (41 to 113 degrees F)	
Relative humidity	20% to 60%		8% to 80%	
Maximum dew point	29 degrees C (84 degrees F)		28 degrees C (82 degrees F)	
Operating voltage	100 to 127 VAC or 200 to 240 VAC	200 to 240 VAC	n/a	
Operating frequency	50 to 60 +/- 3 Hz		n/a	
Power consumption	650 watts maximum	1100 watts maximum	n/a	
Power source loading	0.663 kVA maximum	1.122 kVA maximum	n/a	
Thermal output	2218 BTU/hour maximum	3754 BTU/hour maximum	n/a	

Power 710 and Power 730 operating environment				
Description	**Operating**		**Non-operating**	
	Power 710	**Power 730**	**Power 710**	**Power 730**
Maximum altitude	3048 m (10,000 ft)		n/a	
Noise level reference point (operating/idle)	6.0 bels	6.3 bels	n/a	

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 will take 2U (2 EIA units) of rack space.

Table 1-2 Physical dimensions

Dimension	Power 710 (Model 8231-E2B)	Power 730 (Model 8231-E2B)
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	5.7 kg (12.6 lb)	5.7 kg (12.6 lb)

Figure 1-2 shows the rear view of a Power 730 system.

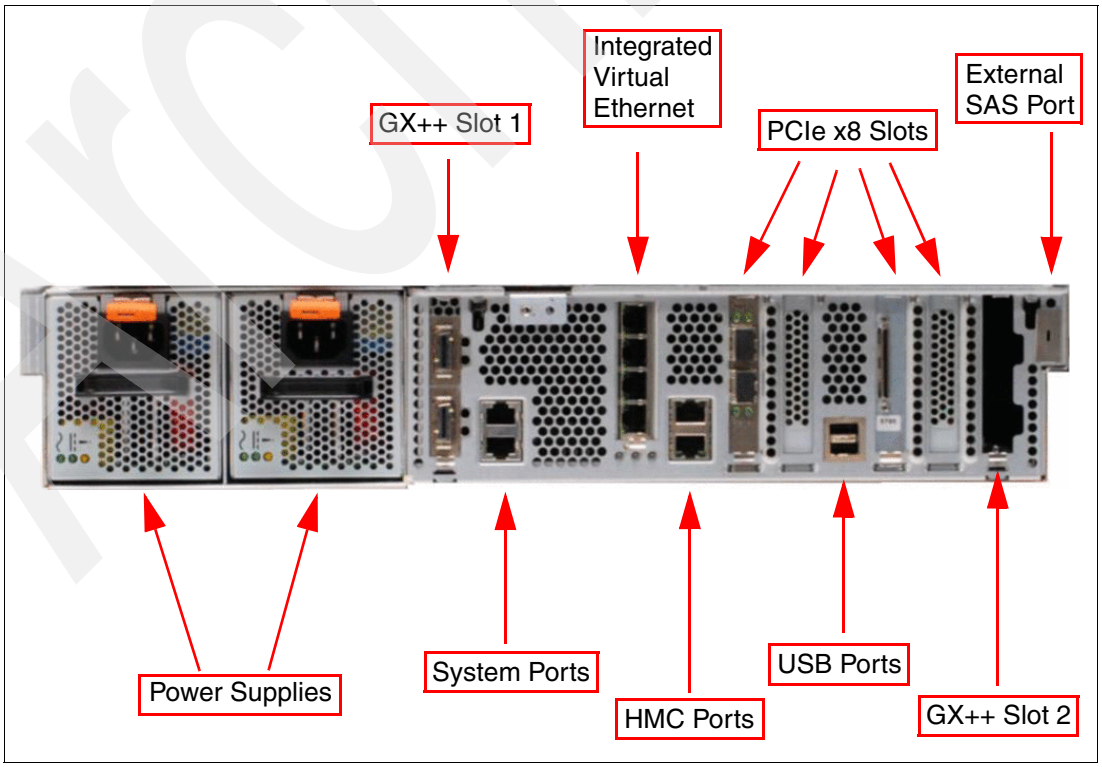


Figure 1-2 Rear view of a Power 730 system

1.4 System features

The system chassis contains one processor module (Power 710) or two processor modules (Power 730). Each POWER7 processor module has either 4-cores, 6-cores, or 8-cores. 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 32 MB of L3 cache (4 MB per core).

1.4.1 Power 710 system features

The following summary describes the standard features:

- Rack-mount (2U) chassis
- Single processor module:
  - 4-core configuration using one 4-core 3.0 GHz processor modules
  - 6-core configuration using one 6-core 3.7 GHz processor modules
  - 8-core configuration using one 8-core 3.55 GHz processor modules
- Up to 64 GB of 1066 MHz DDR3 ECC memory
- Choice of three disk/media backplanes:
  - Six 2.5-inch HDD/SSD and one DVD bay
  - Six 2.5-inch HDD/SSD with Dual Write Cache RAID and with an external SAS port
  - Three 2.5-inch HDD/SSD/Media backplane with one tape drive bay and one DVD bay
- Choice of three Integrated Virtual Ethernet daughter cards:
  - Quad-port 1 Gb Copper IVE
  - Dual-port 10 Gb SFP+ Twinax Cu. IVE
  - Dual-port 10 Gb SFP+ Fiber IVE
- Four PCIe x8 low profile slots
- One GX++ slot used for server clustering configurations
- Integrated:
  - Service Processor
  - EnergyScale technology
  - Hot-swap and redundant cooling
  - Three USB ports; two system ports
  - Two HMC ports
- Optional redundant, 1725 Watt ac hot-swap power supplies

1.4.2 Power 730 system features

The following summary describes the standard features:

- Rack-mount (2U) chassis
- Two processor modules:
  - 8-core configuration using two 4-core 3.0 GHz processor modules
  - 8-core configuration using two 4-core 3.7 GHz processor modules
  - 12-core configuration using two 6-core 3.7 GHz processor modules
  - 16-core configuration using two 8-core 3.55 GHz processor modules
- Up to 128 GB of 1066 MHz DDR3 ECC memory

- Choice of three disk/media backplanes:
  - Six x 2.5-inch HDD/SSD and one DVD bay
  - Six x 2.5-inch HDD/SSD with Dual Write Cache RAID and with an external SAS port
  - Three x 2.5-inch HDD/SSD/Media backplane with one tape drive bay and one DVD bay
- Choice of three Integrated Virtual Ethernet (IVE) daughter cards:
  - Quad-port 1 Gb Copper IVE
  - Dual-port 10 Gb SFP+ TwinAx Cu. IVE
  - Dual-port 10 Gb SFP+ Fiber IVE
- Four PCIe x8 low profile slots
- Two GX++ slots used for server clustering configurations
- Integrated:
  - Service Processor
  - EnergyScale technology
  - Hot-swap and redundant cooling
  - Three USB ports; two system ports
  - Two HMC ports
- Two power supplies, 1725 Watt ac, hot-swap

1.4.3 Minimum features

Each system has a minimum feature-set in order to be valid.

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 (#2145), the initial order must also include one additional HDD/SSD, Mirrored System Disk Level Specify Code, and a System Console Indicator. A DVD is defaulted on every order but can be de-selected. A DVD-ROM or DVD-RAM must be accessible by the system.

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

1.4.4 Power supply features

One 1725 watt ac power supply (#5603) is required for the Power 710. A second power supply is optional. Two 1725 watt ac power supplies (#5603) 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 of the processor modules in the system houses a single POWER7 processor chip. The processor chip has either 4-cores, 6-cores, or 8-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 codes features.

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

Table 1-3 Processor features for the Power 710

Feature code	Processor module description
#8350	4-core 3.0 GHz POWER7 processor module
#8349	6-core 3.7 GHz POWER7 processor module
#8359	8-core 3.55 GHz POWER7 processor module

The Power 730 requires that two identical processor modules be installed. The available processor features can be seen in Table 1-4.

Table 1-4 Processor features for the Power 730

Feature code	Processor module description
#8350	4-core 3.0 GHz POWER7 processor module
#8348	4-core 3.7 GHz POWER7 processor module
#8349	6-core 3.7 GHz POWER7 processor module
#8359	8-core 3.55 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 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 in the configurator. One additional memory riser card, feature #5265, can be installed in the Power 710. Three additional memory riser cards, feature #5265, can be installed in the Power 730. Each memory riser card provides four DDR3 DIMM slots. DIMMs are available in capacities of 2 GB, 4 GB and 8 GB at 1066 MHz. DIMMs are installed in pairs.

Table 1-5 shows memory features available on the systems.

Table 1-5 Summary of memory features

Feature code	Feature capacity	Access rate	DIMMs
#4525	4GB	1066 MHz	2 x 2 GB DIMMs
#4526	8 GB	1066 MHz	2 x 4 GB DIMMs
#4527	16 GB	1066 MHz	2 x 8 GB DIMMs

It is generally best that memory 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. However, balancing memory fairly evenly across multiple memory riser cards, as compared to balancing memory exactly evenly, typically has a very small performance difference.

1.5 Disk and media features

Each system features one SAS DASD controller with three storage backplane options. Table 1-6 shows the available disk drive feature codes that can be installed in the in the Power 710 and Power 730.

Table 1-6 Disk drive feature code description

Feature code	Description	OS support
#1775	177 GB SFF-1 SSD w/ eMLC	AIX, Linux
#1787	177 GB SFF-1 SSD w/ eMLC	IBM i
#1790	600 GB 10K RPM SAS SFF Disk Drive	AIX, Linux
#1885	300 GB 10K RPM SFF SAS Disk Drive	AIX, Linux
#1886	146.8 GB 15K RPM SFF SAS Disk Drive	AIX, Linux
#1888	139.5 GB 15K RPM SFF SAS Disk Drive	IBM i
#1911	283 GB 10K RPM SFF SAS Disk Drive	IBM i
#1916	571 GB 10k RPM SAS SFF Disk Drive	IBM i
#1917	146 GB 15k RPM SAS SFF-2 Disk Drive	AIX, Linux
#1925	300GB 10k RPM SAS SFF-2 Disk Drive	AIX, Linux
#1947	139 GB 15k RPM SAS SFF-2 Disk Drive	IBM i
#1956	283 GB 10k RPM SAS SFF-2 Disk Drive	IBM i
#1962	571 GB 10k RPM SAS SFF-2 Disk Drive	IBM i
#1964	600 GB 10k RPM SAS SFF-2 Disk Drive	AIX, Linux
#1995	177 GB 1.8” SATA Solid State Drive	AIX, Linux
#1996	177 GB 1.8” SATA Solid State Drive	IBM i

Table 1-7 shows the available disk drive feature codes to be installed in an I/O enclosure external to the Power 710 and Power 730.

Table 1-7 Non CEC Disk drive feature code description

Feature code	Description	OS support
#3647	146 GB 15k RPM SAS Disk Drive	AIX, Linux
#3648	300 GB 15k RPM SAS Disk Drive	AIX, Linux
#3649	450 GB 15k RPM SAS Disk Drive	AIX, Linux
#3658	428 GB 15k RPM SAS Disk Drive	IBM i

Feature code	Description	OS support
#3677	139.5 GB 15k RPM SAS Disk Drive	IBM i
#3678	283.7 GB 15k RPM SAS Disk Drive	IBM i

Note: Be aware of the following considerations for disks:

- SAS-bay-based SSD options are enhanced with a 177 GB SSD with eMLC (#1775, #1787), which provides 2.5 times more capacity per drive than the current 69 GB SSD (#1890 and #1909). The 177 GB drive provides a much improved cost per gigabyte and requires a smaller number of SAS bays for the same number of gigabytes.
- The disk features #1775 and #1787 require the following considerations:
  - SFF features #1775 and #1787 are supported in the internal disk bays of the Power 710 and Power 730.
  - SSDs and HDDs are not allowed to mirror each other.
  - SSDs are not supported by PCIe LP 2-x4-port SAS adapter 3 Gb (#5278).
  - An external EXP 12S SAS drawer (# 5886) containing SSD drives cannot be attached to the external SAS port on the Power 710 or Power 730.
  - If IBM i is the primary operating system (#2145), a minimum of two HDDs/SSDs are required.
  - No internal HDD/SSD is required if feature #0837 (Boot from SAN) is selected.In this case, a Fibre Channel or Fibre Channel over Ethernet adapter must also be ordered.
  - In a Power 710 and Power 730, SSDs and HDDs cannot be mixed in storage backplane #5263. SSDs and HDDs can be mixed with storage backplane feature #5267 or #5268. However, they cannot be in the same RAID array.
- The disk features #1995 and #1996 require a PCIe LP RAID & SSD SAS adapter 3 Gb (#2053). For more detailed information about this option, see 2.10.4, “PCIe RAID & SSD SAS adapter” on page 44.
- SSF-2 disk drives require a Gen2 SFF carrier/tray for use in an EXP 24S I/O drawer(#5887)

If you need more disks than available with the internal disk bays, you can attach additional external disk subsystems.

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

For more detailed information about the available external disk subsystems, see 2.13, “External disk subsystems” on page 52.

The Power 710 and Power 730 have a slim media bay that can contain an optional DVD-RAM (#5762) and a tape bay (only available with #5263) that can contain a tape drive or removable disk drive. The Power 710 has a slimline media bay, and a half-high bay that can contain an optional tape drive or removable disk drive equipped with USB Internal Docking Station for Removable Disk Drive (#1103) drive or external docking station (#1104).

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
#1124	DAT160 80/160 GB SAS TAPE DRIVE
#1123	Internal Docking Station for Removable Disk Drive
#1106	USB 160 GB Removable Disk Drive
#1107	USB 500 GB Removable Disk Drive

For more detailed information about the internal disk features, see 2.11, “Internal storage” on page 47.

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

The Power 710 and Power 730 servers do not support the attachment of any I/O drawers.

1.7 Comparison between models

The Power 710 and Power 730 offer a variety of configuration options where the POWER7 processor has 4-cores at 3.0 GHz, 6-cores at 3.7 GHz, or 8-cores at 3.55 GHz.

The POWER7 processor has 4 MB of on-chip L3 cache per core. For the 4-core version there is 16 MB, for the 6-core version there is 24 MB of L3 cache available, whereas for the 8-core version there is 32 MB of L3 cache available.

Table 1-9 summarizes the various processor core options and frequencies and matches them to the L3 cache sizes.

Table 1-9 Summary of processor core counts, core frequencies, and L3 cache sizes

System	Cores per POWER7 chip	Frequency (GHz)	L3 Cache per POWER7 chip	Min/Max cores per system
Power 710	4	3.0	16 MB	4
Power 710	6	3.7	24 MB	6
Power 710	8	3.55	32 MB	8
Power 730	4	3.0	16 MB	8
Power 730	6	3.7	24 MB	12
Power 730	8	3.55	32 MB	16

1.8 Build to Order

You can perform a Build to Order or *a la carte* configuration 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 Editions. These solutions are available at initial system order time with a starting configuration that is ready to run as is.

1.9 IBM Editions

IBM Editions are available only as 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/size of disk, SSD, Fibre Channel adapters, or Fibre Channel over Ethernet (FCoE) adapters shipped with the server are the only features that determine if a customer is entitled to a processor activation at no additional charge.

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 AIX 5.3 or 6.1, or 7.1, IBM i 6.1.1 or IBM i 7.1, and SUSE Linux Enterprise Server or Red Hat Enterprise Linux. If you choose AIX 5.3, 6.1 or 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 or HDD/SSD/Fibre Channel/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 Editions:

- A minimum of 1 GB memory per core on the 4-core Power 710 (#8350) , 2 GB memory per core on the 6-and 8-core Power 710 (#8349 and #8359), and 4 GB memory per core on the Power 730 is needed to qualify for the IBM Edition, except on the 6-core IBM Edition where there is a 16 GB minimum memory requirement for the Power 710. There can be various valid memory configurations that meet the minimum requirement.
- Also, a minimum of two HDDs, or two SSDs, or two Fibre Channel adapters, or two FCoE adapters are required. You only need to meet one of these disk/SSD/FC/FCoE criteria. Partial criteria cannot be combined.

1.10 Hardware Management Console models

If you want to implement partitions, a Hardware Management Console (HMC) or the Integrated Virtualization Manager (IVM) or the new released IBM Systems Director Management Console (SDMC) is required to manage the servers. Multiple IBM POWER6® and POWER7 processor-based servers can be supported by a single HMC or SDMC.

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

If an HMC is used to manage any POWER7 processor-based server, the HMC must be a rack-mount CR3, or later, or deskside C05, or later.

The HMC provides a set of functions that can be used to manage the system, including these:

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

Several HMC models are supported to manage POWER7 processor-based systems. Four models (7042-C07, 7042-C08, 7042-CR5, and 7042-CR6) are available for ordering at the time of writing, but you can also use one the withdrawn models listed in Table 1-10.

Table 1-10 HMC models supporting 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	Available	IBM 7042 Model C07 Deskside Hardware Management Console
7042-C08	Available	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	Available	IBM 7042 Model CR5 Rack-mounted Hardware Management Console
7042-CR6	Available	IBM 7042 Model CR6 Rack-mounted Hardware Management Console

For the IBM Power 710 and IBM Power 730, the Licensed Machine Code Version 7 Revision 720 is required.

Note: 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, and POWER7 processor-based servers. Licensed Machine Code Version 6 (#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 3 GB of RAM minimum and be a rack-mount CR3 model, or later, or deskside C06, or later.

IBM SDMC

The SDMC is intended to be used in the same manner as the HMC. It provides the same functionality, including hardware, service, and virtualization management, for your Power Systems server and Power Systems blades. Because SDMC uses IBM Systems Director Express® Edition, it also provides all Systems Director Express capabilities, such as monitoring of operating systems and creating event action plans.

Much of the SDMC function is equivalent to the HMC. This includes:

- Server (host) management
- Virtualization management
- Transition of configuration data: no configuration changes are required when a client moves from HMC management to SDMC management.
- Redundancy and high availability: The SDMC offers console redundancy similar to the HMC

In most cases, the scalability and performance of SDMC matches that of a current HMC. This includes both the number of systems (hosts) and the number of partitions (virtual servers) that can be managed. Currently, 48 small-tier entry servers or 32 large-tier servers can be managed by the SDMC with up to 1,024 partitions (virtual servers) configured across those managed systems (hosts).

The SDMC can be obtained as a hardware appliance-in the same manner as an HMC. Hardware appliances support managing all Power Systems servers. The SDMC can optionally be obtained in a virtual appliance format, capable of running on VMware (ESX/i 4, or later), and KVM (Red Hat Enterprise Linux® (RHEL) 5.5). The virtual appliance is only supported or managing small-tier Power® servers and Power Systems blades.

Table 1-11 and Table 1-12 detail whether the SDMC software appliance, hardware appliance, or both are supported for each model.

Table 1-11 Type of SDMC appliance support for POWER7 based server

POWER7 models	Type of SDMC appliance supported
7891-73X, 7891-74X PS703, PS704	Hardware or software appliance
8202-E4B (IBM Power 720 Express®)	Hardware or software appliance
8205-E6B (IBM Power 740 Express)	Hardware or software appliance
8406-70Y PS700	Hardware or software appliance
8406-71Y PS701,702	Hardware or software appliance
8231-E2B (IBM Power 710 nd IBM Power 730 Express)	Hardware or software appliance
8233-E8B (IBM Power 750 Express)	Hardware or software appliance
8236-E8C (IBM Power 755)	Hardware or software appliance
9117-MMB (IBM Power 770)	Hardware appliance only
9119-FHB (IBM Power 795)	Hardware appliance only
9179-MHB (IBM Power 780)	Hardware appliance only

Table 1-12 Type of SDMC appliance support for POWER6 based server

POWER6 models	Type of SDMC appliance supported
7778-23x JS23	Hardware or software appliance
7998-60X JS12	Hardware or software appliance
7998-61X JS22	Hardware or software appliance
8203-E4A (Power 520 Express)	Hardware or software appliance
8204-E8A (Power 550 Express)	Hardware or software appliance
8234-EMA (Power 560 Express)	Hardware appliance only
8261-E4S (IBM Smart Cubes)	Hardware or software appliance
9406-MMA (Power 570)	Hardware appliance only
9407-M15 (Power 520 Express)	Hardware or software appliance
9408-M25 (Power 520 Express)	Hardware or software appliance
9409-M50 (Power 550 Express)	Hardware or software appliance
9117-MMA (Power 570)	Hardware appliance only
9119-FHA (Power 595)	Hardware appliance only
9125-F2A (Power 575)	Hardware appliance only

The IBM SDMC Hardware Appliance requires:

- An IBM 7042 Rack-mounted Hardware Management Console and IBM SDMC indicator (#0963).

Note: When ordering #0963, you must also order features #0031(No Modem) , #1946 (additional 4GB (2X2GB), 1.333GHz DDR3) and #1998 (additional 500GB SATA HDD). Feature # 0963 replaces the HMC software with IBM Systems Director Management Console Hardware Appliance V6.7.3 (5765-MCH).

Neither an external modem (#0032) nor an internal modem (#0033) may be selected with IBM SDMC (#0963).

To run HMC LMC (#0962), you cannot order the additional storage (#1998); however, you can order the additional memory (#1946) if wanted.

The IBM SDMC Virtual Appliance requires:

- IBM Systems Director Management Console V6.7.3 (5765-MCV)

Note: For the software appliance, customer is responsible for providing the hardware and and virtualization environment.

At a minimum, the following resources should be available to the virtual machine:

- 2.53 GHz Intel Xeon E5630, Quad Core processor
- 500 MB storage
- 6 GB memory

The hypervisors supported are:

- VMware (ESX/i 4.0, or later)
- KVM (RHEL 5.5)

SDMC on Power Systems POWER6 processor-based servers and blades requires eFirmware level 3.5.7. SDMC on Power Systems POWER7 processor-based servers and blades requires eFirmware level 7.3.0.

1.11 System racks

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

Note: 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 has to integrate the newly ordered external disk drawer in a 19-inch rack before shipping the MES order.

If a system is to be installed in a non-IBM rack or cabinet, ensure that the rack meets the requirements described in 1.11.10, “OEM rack” on page 21.

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

1.11.1 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 that will accommodate #7188 PDUs. All PDUs required for application in these racks must be installed horizontally in the rear of the

rack. Each horizontally mounted PDU occupies 1U of space in the rack, and therefore reduces the space available for mounting servers and other components.

1.11.2 IBM 7014 Model T00 rack

The 1.8 Meter (71-in.) Model T00 is compatible with past and present IBM Power systems. The T00 rack has the following features:

- 36 EIA units (36 U) of usable space.
- Optional removable side panels.
- Optional highly perforated front door.
- Optional side-to-side mounting hardware for joining multiple racks.
- Standard business black or optional white color in OEM format.
- Increased power distribution and weight capacity.
- Support for both ac and dc configurations.
- The rack height is increased to 1926 mm (75.8 in.) if a power distribution panel is fixed to the top of the rack.
- Up to four power distribution units (PDUs) can be mounted in the PDU bays (see Figure 1-3 on page 17), but others can fit inside the rack. See 1.11.7, “The ac power distribution unit and rack content” on page 17.
- Weights:
  - T00 base empty rack: 244 kg (535 lb.)
  - T00 full rack: 816 kg (1795 lb.)

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 differ in the Model T42 rack from the Model T00:

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

1.11.4 Feature code 0555 rack

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

1.11.5 Feature code 0551 rack

The 1.8 Meter Rack (#0551) is a 36 EIA unit rack. The rack that is delivered as #0551 is the same rack 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 #0551.

1.11.6 Feature code 0553 rack

The 2.0 Meter Rack (#0553) is a 42 EIA unit rack. The rack that is delivered as #0553 is the same rack 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 #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 ac power distribution units #9188 and #7188 and ac Intelligent PDU+ #5889 and #7109.

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

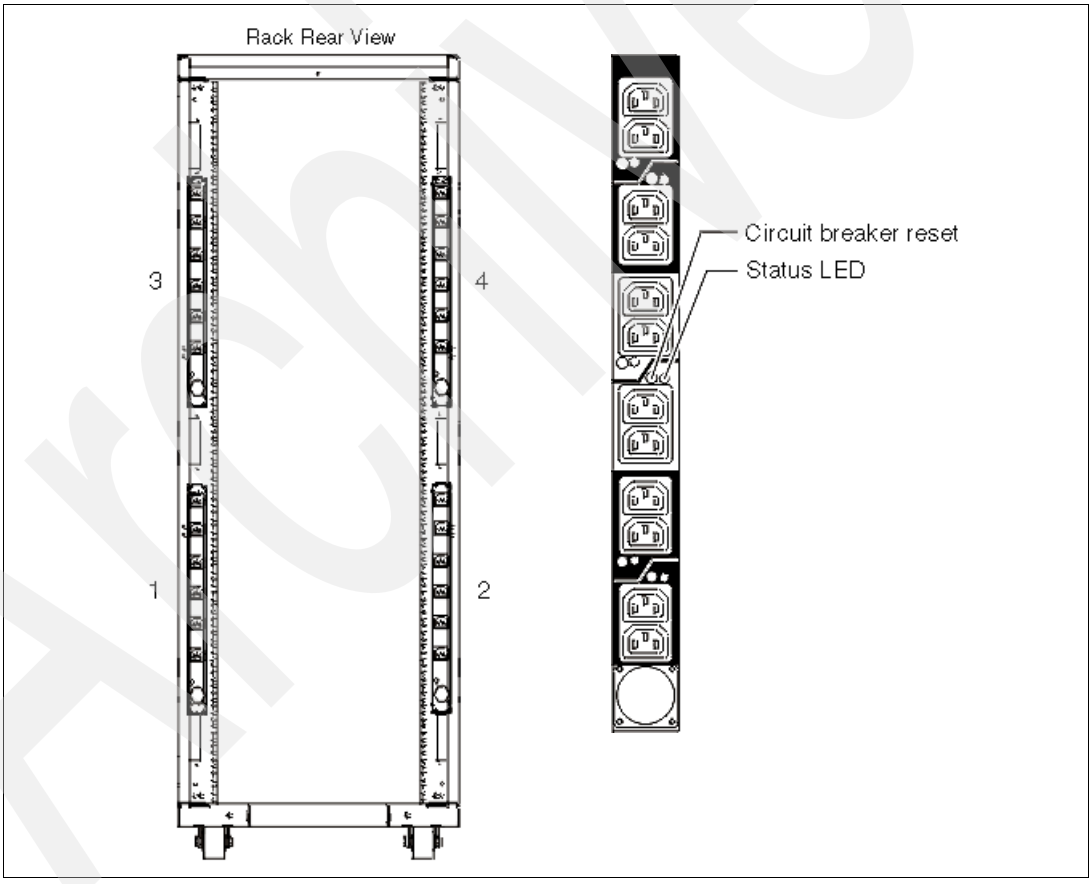


Figure 1-3 PDU placement and PDU view

For detailed power cord requirements and power cord feature codes, see the IBM Power Systems Hardware Information Center at the following website:

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

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

The Intelligent PDU+, base option, 1 EIA Unit, Universal, UTG0247 Connector (#5889); the Base/Side Mount Universal PDU (#9188); the optional, additional, Universal PDU (#7188); and the Intelligent PDU+ options (#7109) support a wide range of country requirements and electrical power specifications. The #5889 and #7109 PDUs are identical to the #9188 and #7188 PDUs but are equipped with one Ethernet port, one console serial port, and one rs232 serial port for power monitoring.

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.

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

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

The Universal PDUs are compatible with previous models.

Note: Each system drawer to be mounted in the rack requires two power cords, which are not included in the base order. For maximum availability, it is definitely best to connect power cords from the same system to two separate PDUs in the rack, and to connect each PDU to independent power sources.

1.11.8 Rack-mounting rules

Follow these primary rules when mounting the system into a rack:

- The system is designed to be placed at any location in the rack. For rack stability, it is advisable to start filling a rack from the bottom.
- Any remaining space in the rack can be used to install other systems or peripherals, provided that the maximum permissible weight of the rack is not exceeded and the installation rules for these devices are followed.
- Before placing the system into the service position, it is essential that the rack manufacturer's safety instructions have been followed regarding rack stability.

1.11.9 Useful rack additions

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

IBM System Storage 7214 Tape and DVD Enclosure

The IBM System Storage® 7214 Tape and DVD Enclosure 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.

The two bays of the IBM System Storage 7214 Tape and DVD Enclosure can accommodate the following tape or DVD drives for IBM Power servers:

- DAT72 36 GB Tape Drive - up to two drives
- DAT72 36 GB Tape Drive - up to two drives
- DAT160 80 GB Tape Drive - up to two drives
- Half-high LTO Ultrium 4 800 GB Tape Drive - up to two drives
- DVD-RAM Optical Drive - up to two drives
- DVD-ROM Optical Drive - up to two drives

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.

The following optional drive technologies are available for the 7216-1U2:

- DAT160 80 GB SAS Tape Drive (#5619
- DAT320 160 GB SAS Tape Drive (#1402)
- DAT320 160 GB USB Tape Drive (#5673)
- Half-high LTO Ultrium 5 1.5 TB SAS Tape Drive (#8247)
- DVD-RAM - 9.4 GB SAS Slim Optical Drive (#1420 and #1422)
- RDX Removable Disk Drive Docking Station (#1103)

Note: DAT320 160 GB SAS Tape Drive (#1402)and DAT320 160 GB USB Tape Drive (#5673) No Longer Available as of July 15, 2011

In order 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 LP 2-x4-port SAS adapter 3 Gb (#5278) must be installed in the Power 710 and Power 730 server in order 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 (#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 LP 2-x4-port SAS adapter 3 Gb (#5278).

- Attachment by a USB adapter:

  The Removable RDX HDD Docking Station features on 7216 only support 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 (# 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 Half-high LTO Ultrium 5 SAS) with second bay empty
- Two tape drives (DAT160 SAS or Half-high LTO Ultrium 5 SAS) in any combination
- One tape drive (DAT160 SAS or Half-high LTO Ultrium 5 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 Half-high LTO Ultrium 5 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 right bay
- Two RDX Removable HDD Docking Stations

Figure 1-4 shows the 7216 Multi-Media Enclosure.

Figure 1-4 7216 Multi-Media Enclosure

For a current list of host software versions and release levels that support the IBM System Storage 7216 Multi-Media Enclosure, refer to the following 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 consisting 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 screen TFT monitor with truly accurate images and virtually no distortion
- Ability to mount the IBM Travel Keyboard in the 7316-TF3 rack keyboard tray

- Support for 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 key points mentioned are as follows:

- The front rack opening must be 451 mm wide ± 0.75 mm (17.75 in. ± 0.03 in.), and the rail-mounting holes must be 465 mm ± 0.8 mm (18.3 in. ± 0.03 in.) apart on center (horizontal width between the vertical columns of holes on the two front-mounting flanges and on the two rear-mounting flanges). See Figure 1-5 for a top view showing the specification dimensions.

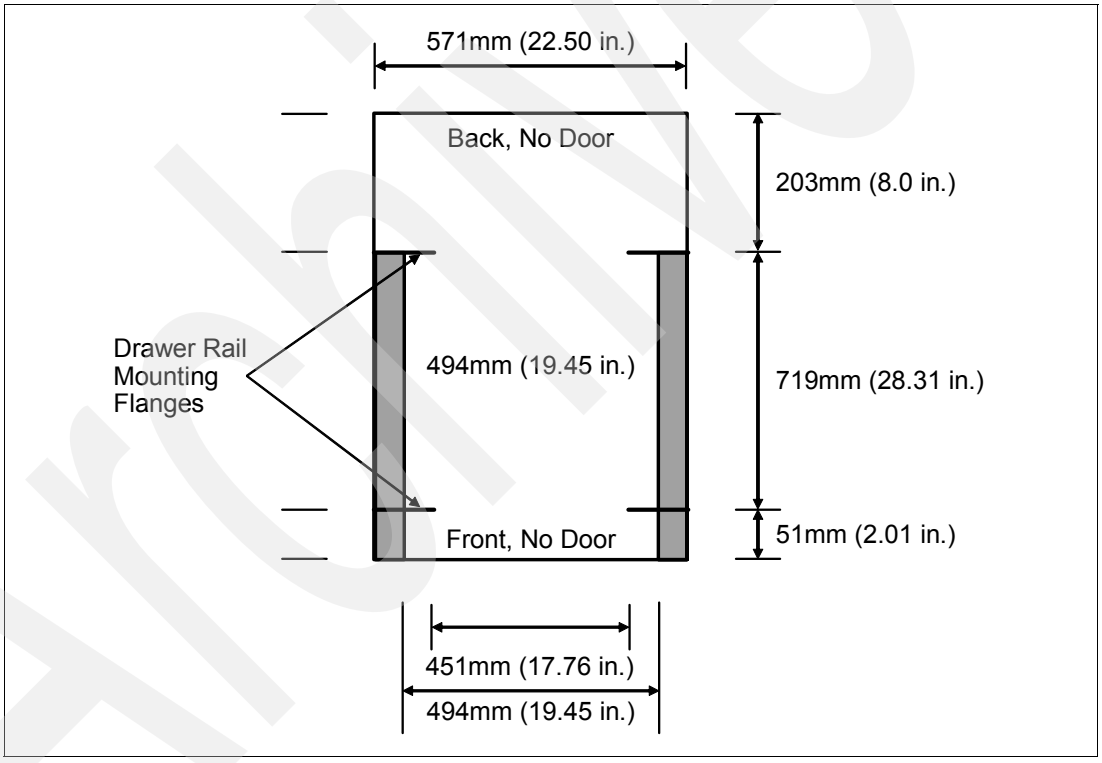


Figure 1-5 Top view of non-IBM rack specification dimensions

- The vertical distance between the mounting holes must consist of sets of three holes spaced (from bottom to top) 15.9 mm (0.625 in.), 15.9 mm (0.625 in.), and 12.67 mm (0.5 in.) on center, making each three-hole set of vertical hole spacing 44.45 mm (1.75 in.) apart on center. Rail-mounting holes must be 7.1 mm ± 0.1 mm (0.28 in. ± 0.004 in.) in diameter. See Figure 1-6 for the top front specification dimensions.

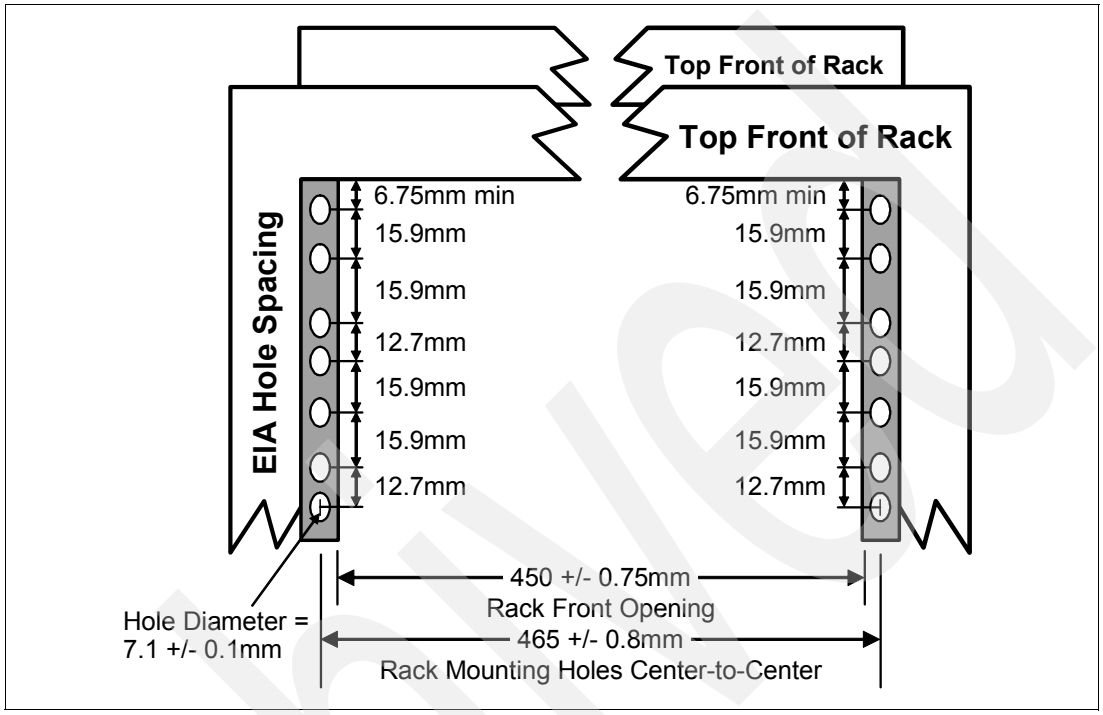


Figure 1-6 Rack specification dimensions, top front view

2

Architecture and technical overview

This chapter discusses the overall system architecture represented by Figure 2-1, with its major components described in the following sections. 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.

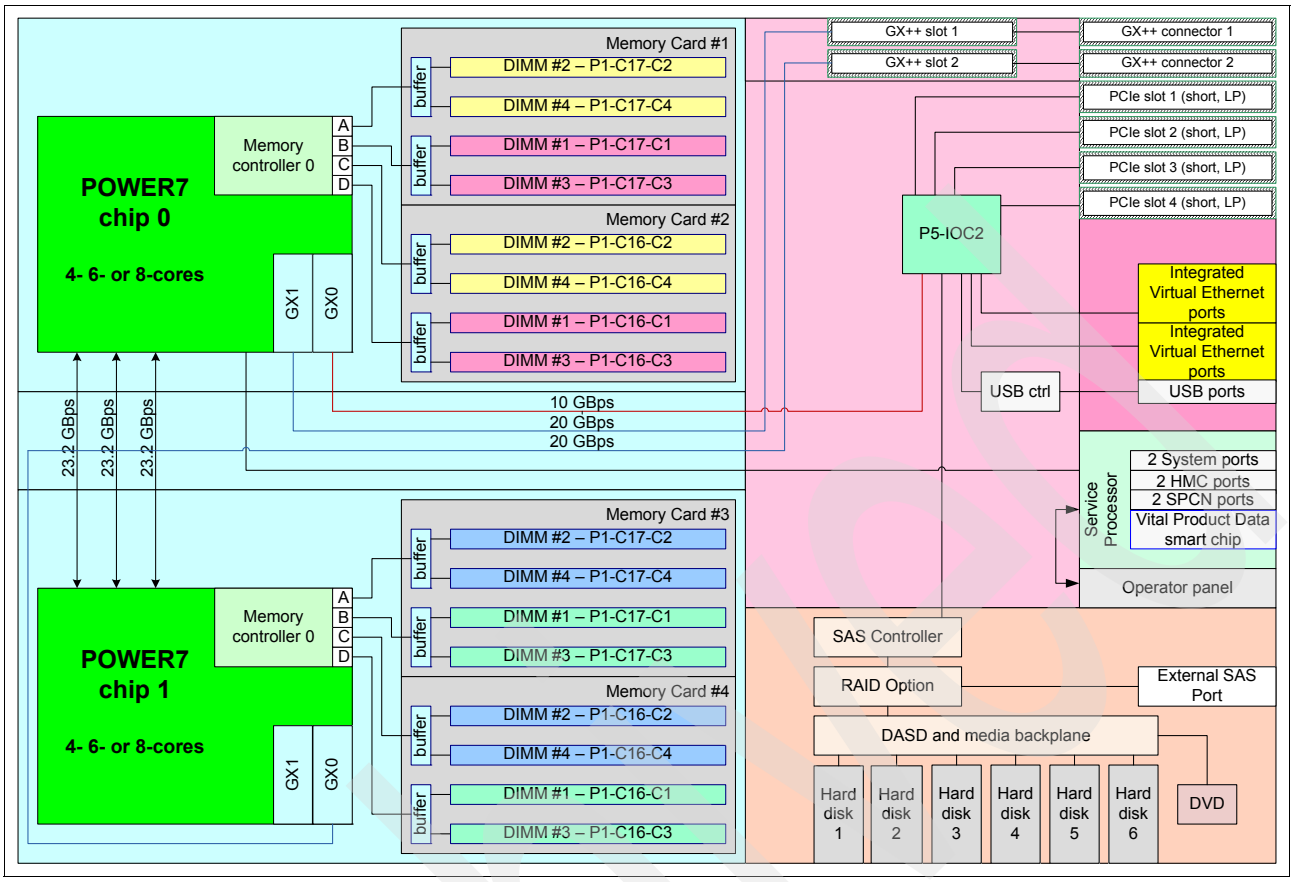


Figure 2-1 Power 730 logic data flow

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 has been matched with innovation across a wide range of related technologies in order to deliver leading throughput, efficiency, scalability, and RAS.

Although the processor is an important component in delivering outstanding servers, many elements and facilities have to be balanced on a server in order to deliver maximum throughput. As with previous generations of systems based on 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.

In many cases, IBM has been innovative in order 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 features:

- 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 signalling
- Exploitation of long-term investment in coherence innovation

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 and IBM POWER6+™ processor-based systems

Figure 2-2 shows the POWER7 processor die layout with the major areas identified; processor cores, L2 cache, L3 cache and chip interconnection, Symmetric Multi Processing (SMP) links, and memory controllers.

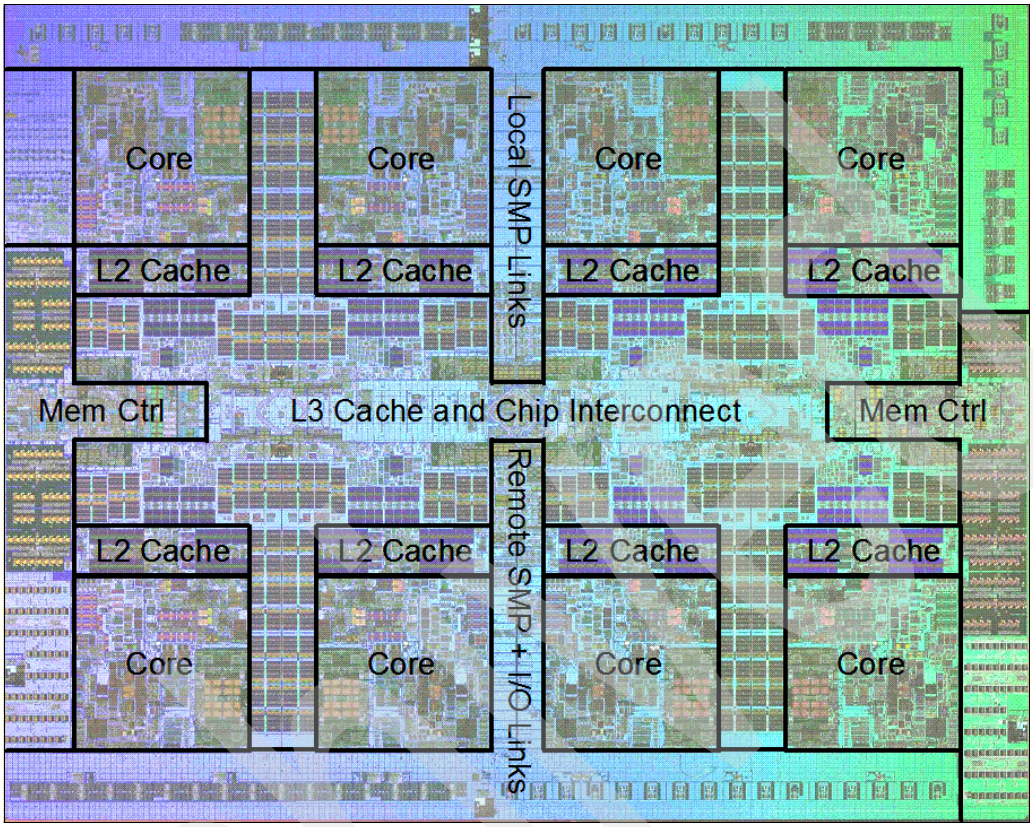


Figure 2-2 POWER7 processor die with key areas indicated

2.1.1 POWER7 processor overview

The POWER7 processor chip is fabricated using the IBM 45 nm Silicon-On-Insulator (SOI) technology using copper interconnect and implements an on-chip L3 cache using eDRAM.

The POWER7 processor chip has an area of 567 mm^2 and is built using 1.2 billion components (transistors). Eight processor cores are on the chip, each with 12 execution units, 256 KB of L2 cache, and access to up to 32 MB of shared on-chip L3 cache.

For memory access, the POWER7 processor includes two Double Data Rate 3 (DDR3) memory controllers, each with four memory channels. To be able to scale effectively, the POWER7 processor uses a combination of local and global SMP links with very high coherency bandwidth and takes advantage of the IBM dual-scope broadcast coherence protocol.

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	► 45 nm lithography ► Copper interconnect ► Silicon-on-Insulator ► eDRAM
Components	1.2 billion components/transistors offering the equivalent function of 2.7 billion (for further details, see 2.1.6, “On-chip L3 cache innovation and Intelligent Cache” on page 31)
Processor cores	8
Max execution threads per core/chip	4/32
L2 cache per core/chip	256 KB/2 MB
On-chip L3 cache per core/chip	4 MB/32 MB
DDR3 memory controllers	2
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 twelve execution units:

- 2 fixed point units
- 2 load store units
- 4 double precision floating point units
- 1 vector unit
- 1 branch unit
- 1 condition register unit
- 1 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

An enhancement in the POWER7 processor is the addition of the SMT4 mode to enable four instruction threads to execute simultaneously in each POWER7 processor core. Thus, the instruction thread execution modes of the POWER7 processor are as follows:

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

Maximizing throughput

SMT4 mode enables the POWER7 processor to maximize the throughput of the processor core by offering an increase in core efficiency. SMT4 mode is the latest step in an evolution of multithreading technologies introduced by IBM. The diagram in Figure 2-3 shows the evolution of simultaneous multithreading.

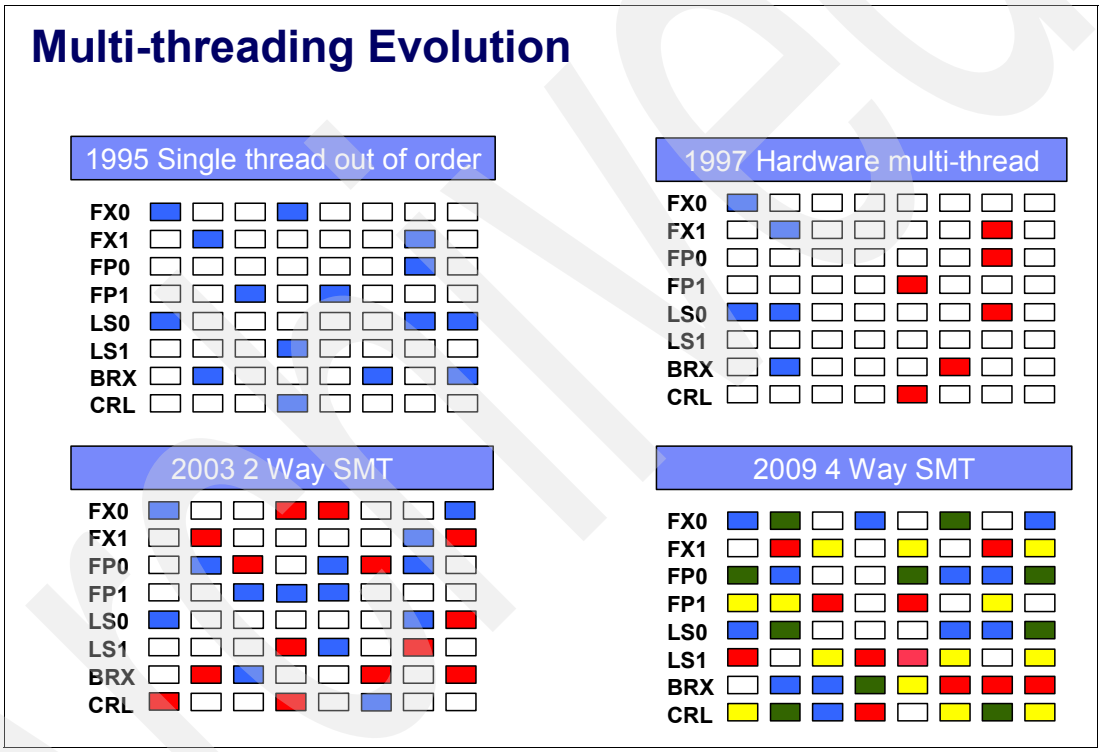


Figure 2-3 Evolution of simultaneous multi-threading

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 very fast individual tasks can get the performance they need for maximum benefit.

2.1.4 Memory access

Each POWER7 processor chip has two DDR3 memory controllers, each with four memory channels (enabling eight memory channels per POWER7 processor chip). Each channel operates at 6.4 GHz and can address up to 32 GB of memory. Thus, each POWER7 processor chip is capable of addressing up to 256 GB of memory.

Figure 2-4 gives a simple overview of the POWER7 processor memory access structure.

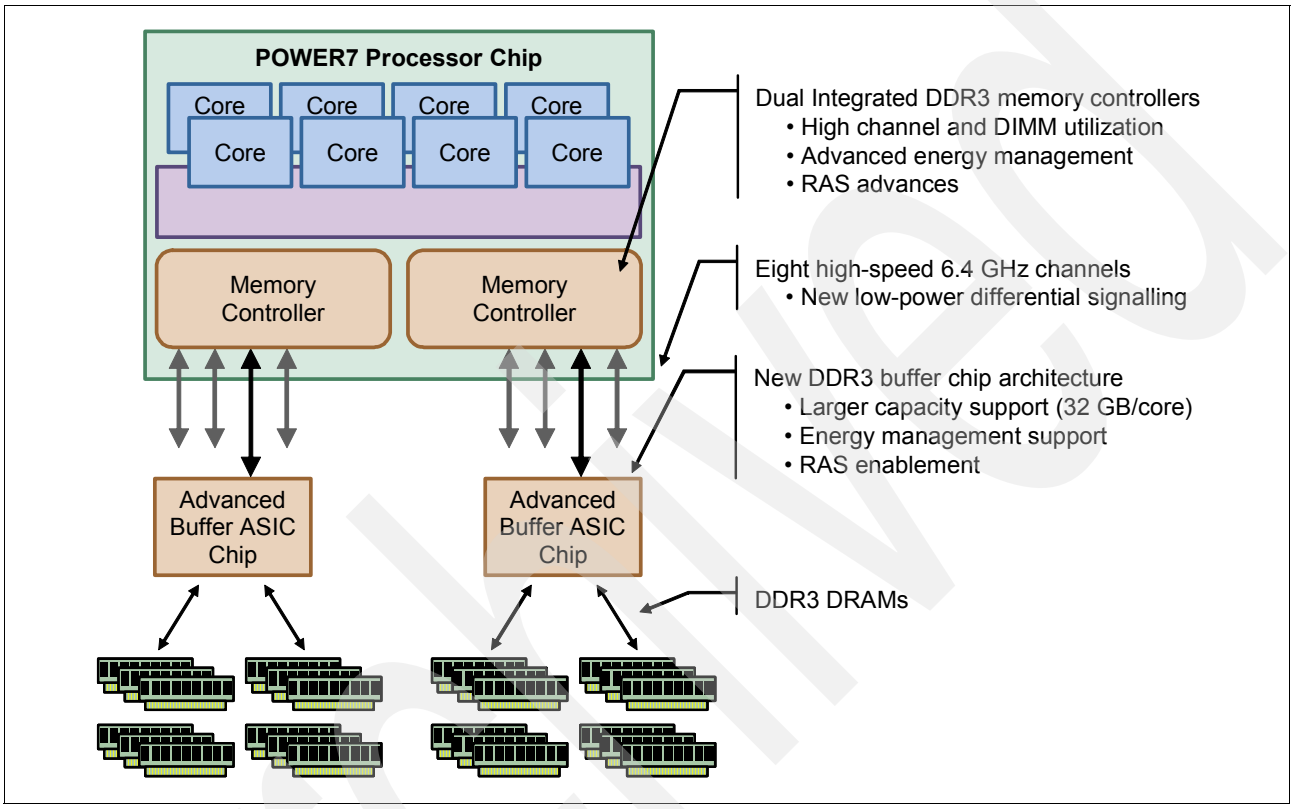


Figure 2-4 Overview of POWER7 memory access structure

2.1.5 Flexible POWER7 processor packaging and offerings

The POWER7 processor forms the basis of a flexible compute platform and can be offered in a number of guises to address differing system requirements.

The POWER7 processor can be offered with a single active memory controller with four channels for servers where higher degrees of memory parallelism are not required.

Similarly, the POWER7 processor can be offered with a variety of SMP bus capacities that are appropriate to the scaling-point of particular server models.

Figure 2-5 outlines the physical packaging options that are supported with POWER7 processors.

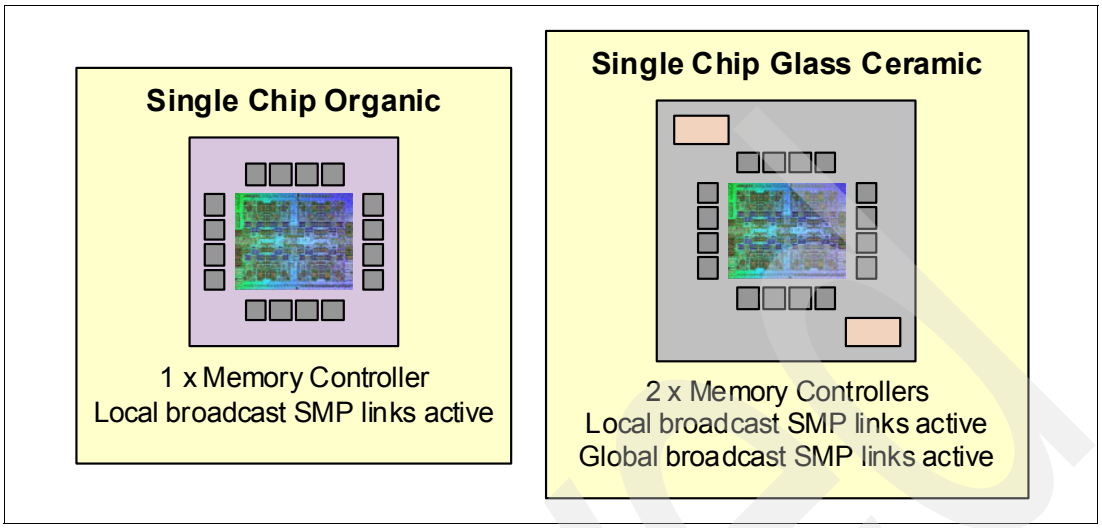


Figure 2-5 Outline of the POWER7 processor physical packaging

POWER7 processors have the unique ability to optimize to various workload types. For example, database workloads typically benefit from very fast processors that handle high transaction rates at high speeds. Web workloads typically benefit more from processors with many threads that allow the breaking down of web requests into many parts and handle them in parallel. POWER7 processors uniquely have the ability to provide leadership performance in either case.

TurboCore mode

Users can opt to run selected servers in TurboCore mode. This mode uses four cores per POWER7 processor chip with access to the entire 32 MB of L3 cache (8 MB per core) and at a faster processor core frequency, which delivers higher performance per core, and might save on software costs for those applications that are licensed per core.

Note: TurboCore is available on the Power 780 and Power 795.

MaxCore mode

MaxCore mode is for workloads that benefit from a higher number of cores and threads handling multiple tasks simultaneously, taking advantage of increased parallelism. MaxCore mode provides up to eight cores and up to 32 threads per POWER7 processor.

POWER7 processor 4-core and 6-core offerings

The base design for the POWER7 processor is an 8-core processor with 32 MB of on-chip L3 cache (4 MB per core). However, the architecture allows for differing numbers of processor cores to be active; 4-cores or 6-cores, as well as the full 8-core version.

In most cases (MaxCore mode), the L3 cache associated with the implementation is dependent on the number of active cores. For a 6-core version, this typically means that 6 x 4 MB (24 MB) of L3 cache is available. Similarly, for a 4-core version, the L3 cache available is 16 MB.

2.1.6 On-chip L3 cache innovation and Intelligent Cache

A breakthrough in material engineering and microprocessor fabrication has 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 signalling 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 the FLR-L3 cache regions for each of the cores on the POWER7 processor die.

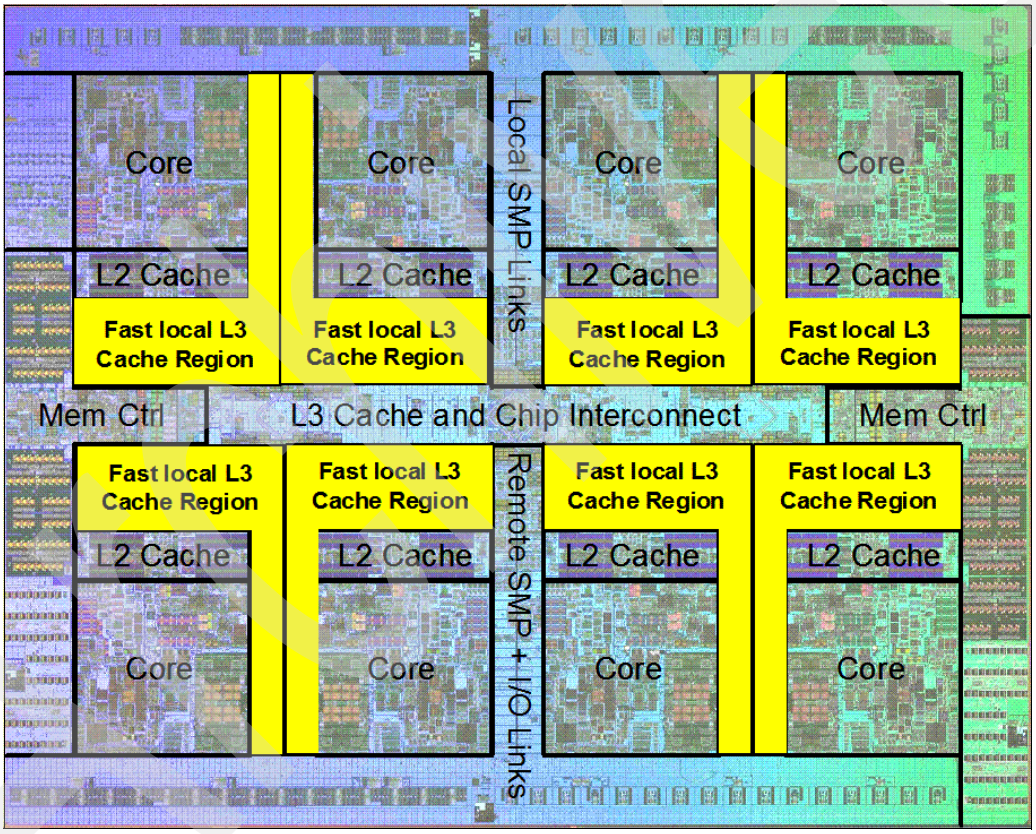


Figure 2-6 Fast local regions of L3 cache on the POWER7 processor

Innovation using eDRAM on the POWER7 processor die is significant for these 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 eDRAM L3 cache requires far less physical space than an equivalent L3 cache implemented with conventional SRAM. IBM on-chip eDRAM uses only a third of the components used in conventional SRAM which has a minimum of 6 transistors to implement a 1-bit memory cell.

- Low energy consumption:

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

2.1.7 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.

2.1.8 Comparison of the POWER7 and POWER6 processors

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

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

Feature	POWER7	POWER6
Technology	45 nm	65 nm
Die size	567 mm^2	341 mm^2
Maximum cores	8	2
Maximum SMT threads per core	4 threads	2 threads
Maximum frequency	4.25 GHz	5 GHz
L2 Cache	256 KB per core	4 MB per core
L3 Cache	4 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	DDR2
I/O Bus	Two GX++	One GX++
Enhanced Cache Mode (TurboCore)	Yes	No
Sleep & Nap Mode[a]	Both	Nap only

a. For more information about Sleep and Nap modes, see 2.18.1, “IBM EnergyScale technology” on page 70

2.2 POWER7 processor modules

The Power 710 and Power 730 server chassis house POWER7 processor modules that host POWER7 processor sockets (SCM) and eight DDR3 memory DIMM slots for each processor module.

The Power 710 server houses one processor module offering 4-core 3.0 GHz, 6-core 3.7 GHz, or 8-core 3.55 GHz configurations.

The Power 730 server houses two processor modules offering 8-core 3.0 and 3.7 GHz, 12-core 3.7 GHz, or 16-core 3.55 GHz configurations.

All off the installed processors must be activated.

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.

Modules and cards

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

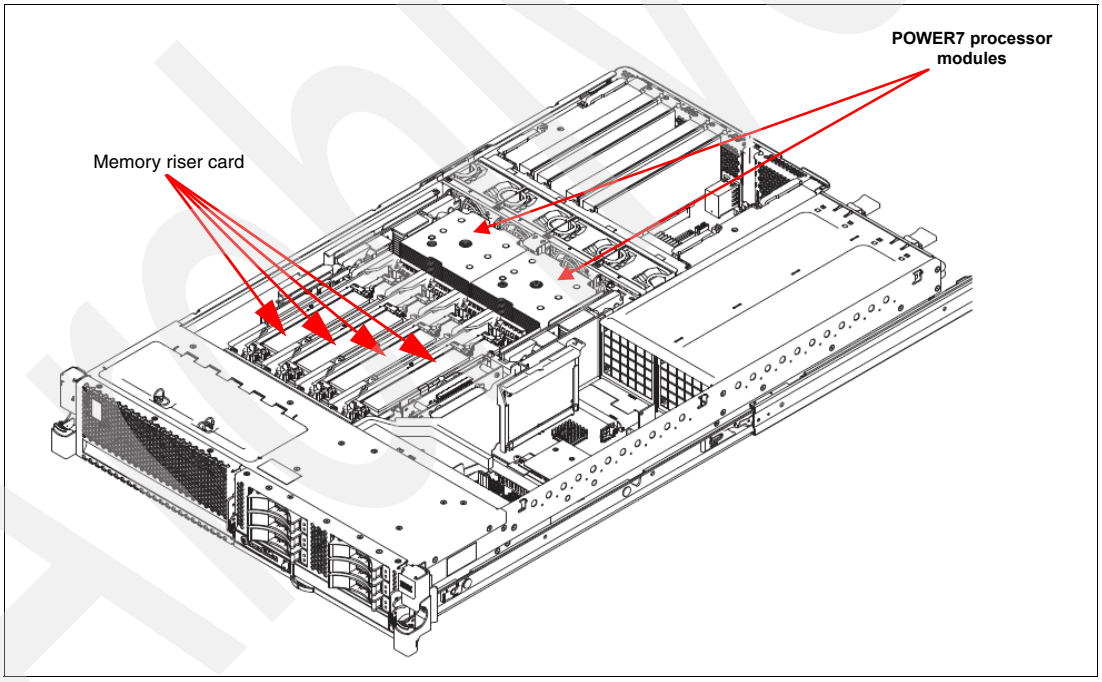


Figure 2-7 Power 730 with two POWER7 processor modules and four memory riser cards

Power 710 and Power 730 systems

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

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

Feature	Cores per POWER7 processor	Frequency (GHz)	Processor activation	Min./Max cores per system	Min./Max processor module
#8350	4	3.0	The 4-core 3.0 GHz requires that four processor activation codes are ordered, available as 4 x #8360 or 2 x #8360 and 2 x #8363	4	1
#8349	6	3.7	The 6-core 3.7 GHz requires that six processor activation codes are ordered, available 6 x #8382 or 3 x #8382 and 3 x #8384	6	1
#8359	8	3.55	The 8-core 3.55 GHz requires that eight processor activation codes are ordered, available 8 x #8372 or 4 x #8372 and 4 x #8375	8	1

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
#8350	4	3.0	The 4-core 3.0 GHz requires that eight processor activation codes are ordered, available as 8 x #8360 or 4 x #8360 and 4 x #8363	8	2
#8348	4	3.7	The 4-core 3.7 GHz requires that eight processor activation codes be ordered, available as 8 x #8381or 4 x #8381 and 4 x #8383) are required	8	2
#8349	6	3.7	The 6-core 3.7 GHz requires that 12 processor activation codes are ordered, available 12 x #8382 or 6 x #8382 and 6 x #8384	12	2
#8359	8	3.55	2 x of the 8-core 3.55 GHz requires that 16 processor activation codes are ordered, available 16 x #8372 or 8 x #8372 and 8 x #8375	16	2

2.3 Memory subsystem

The Power 710 is a one socket system supporting 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. Memory features (two memory DIMMs per feature) supported are 4 GB, 8 GB and 16 GB running at speeds of 1066 MHz. A system with the optional memory riser card installed has a maximum memory of 64 GB.

The Power 730 is a two socket system supporting 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. Memory features (two memory DIMMs per feature) supported are 4 GB, 8 GB and 16 GB run at speeds of 1066 MHz. A system with three optional memory riser cards installed has a maximum memory of 128 GB.

2.3.1 Registered DIMM

Industry standard DDR3 Registered DIMM (RDIMM) technology is used to increase reliability, speed, and density of memory subsystems.

2.3.2 Memory placement rules

The following memory options are orderable:

- 4 GB (2 x 2 GB) Memory DIMMs, 1066 MHz (#4526)
- 8 GB (2 x 4 GB) Memory DIMMs, 1066 MHz (#4526)
- 16 GB (2 x 8 GB) Memory DIMMs, 1066 MHz (#4527)

The minimum DDR3 memory capacity for the system is 8 GB (2 x 4 GB DIMMs). The maximum memory supported is as follows:

- Power 710: 64 GB (four 8 GB DIMMs on each of two memory cards)
- Power 730: 128 GB (four 8 GB DIMMs on each of four memory cards)

Note: DDR2 memory (used in POWER6 processor-based systems) is not supported in POWER7 processor-based systems.

Figure 2-8 shows the physical memory DIMM topology.

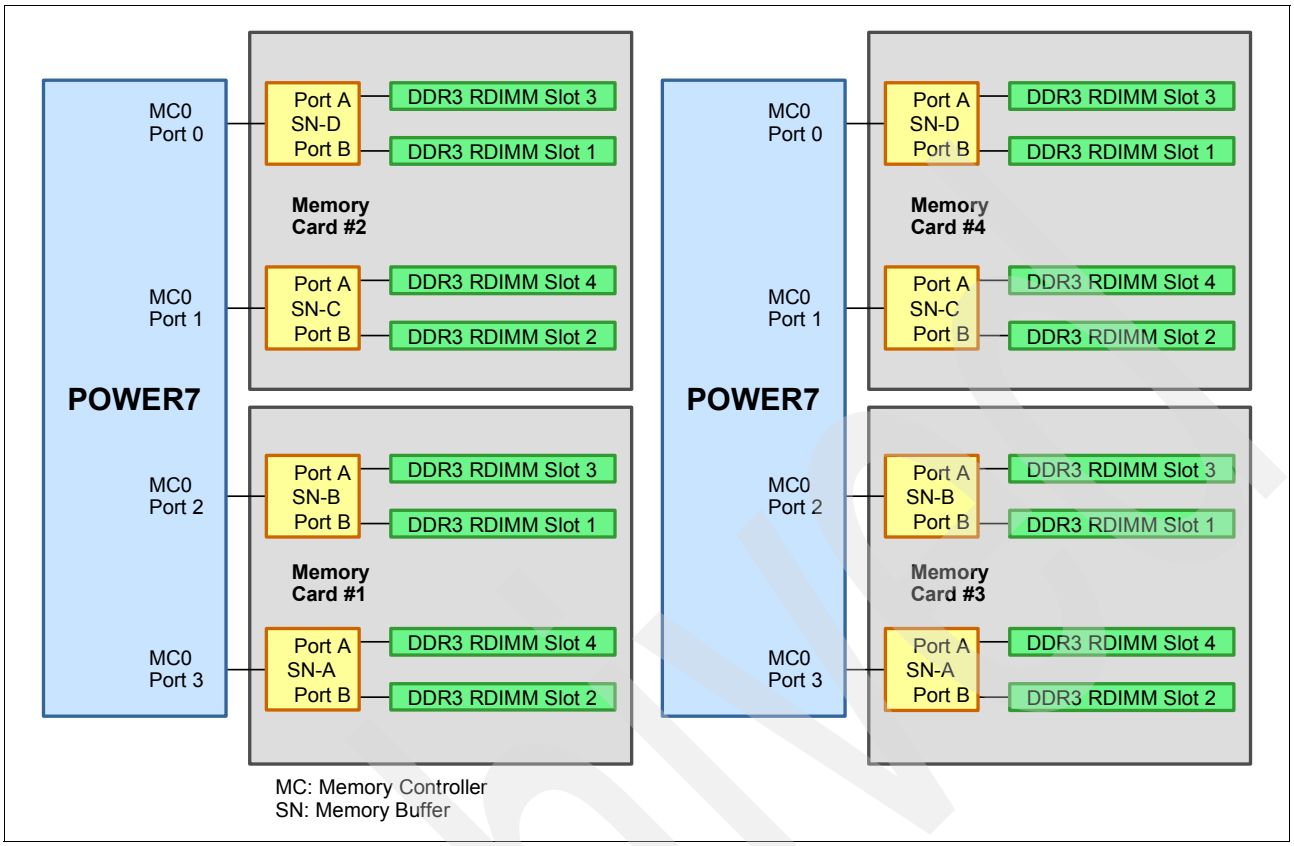


Figure 2-8 Memory DIMM topology for the Power 730

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 #5265) with additional four DIMM sockets. Maximum system memory is 32 GB without feature #5265 and 64 GB with one feature #5265.
- The Power 730 offers three optional memory riser card features (3 x #5265) with an additional four DIMM sockets per feature. Maximum system memory is 32 GB without feature #5265 and 128 GB with three feature #5265.
- A system can be ordered with a single memory feature #4525, #4526 or #4527. When adding additional memory to a system with one feature #4525, a second #4525 must be ordered, and then all memory must be ordered in increments of two #4525. When adding additional memory to a system with one feature #4526, a second #4526 must be ordered, and then all memory must be ordered in increments of two #4526. When adding additional memory to a system with one feature #4527, a second #4527 must be ordered, and then all memory must be ordered in increments of two #4527.
- Feature #4525 can be ordered as a single feature on #8350 processor only, when no additional memory is present or as an MES to pair an exiting #4525 feature on an #8350 processor. At all other times (and always on processors #8349 & #8359) #4525 must be ordered in quantities of 2.
- Feature #4525 #4526 and feature #4527 are not allowed on the same memory riser card. However, memory features can vary from one memory riser card to another.

It is generally best 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 very small performance difference.

Take into account any plans for future memory upgrades when deciding which memory feature size to use at the time of initial system order.

Table 2-5 shows the installation slots for memory DIMMs for two POWER7 modules and four memory cards.

Table 2-5 Memory DIMM installation sequence and slots

DIMM-pair	Memory card	DIMM slots
First	One	► Slot 1 (P1-C17-C1) ► Slot 2 (P1-C17-C2))
Second	One	► Slot 3 (P1-C17-C3) ► Slot 4 (P1-C17-C4)
Third Fourth	Three	► Slot 1 (P1-C15-C1) ► Slot 2 (P1-C15-C2) ► Slot 3 (P1-C15-C3) ► Slot 4 (P1-C15-C4)
Fifth Sixth	Two	► Slot 1 (P1-C16-C1) ► Slot 2 (P1-C16-C2) ► Slot 3 (P1-C16-C3) ► Slot 4 (P1-C16-C4)
Seventh Eighth	Four	► Slot 1 (P1-C14-C1) ► Slot 2 (P1-C14-C2) ► Slot 3 (P1-C14-C3) ► Slot 4 (P1-C14-C4)
Notes: ► Only the4 GB Feature #4525 (2 x 2 GB DIMM) can be installed as a single feature in DIMM pairs (default for #8350 processor). All subsequent memory features must be installed in quads (2 x feature). ► Only the 8 GB Feature #4526 (2 x 4 GB DIMM) can be installed as a single feature in DIMM pairs (default for #8349 and #8359 processors). All subsequent memory features must be installed in quads (2 x feature) ► DIMMs on a given memory card can vary from DIMMs on other memory cards. ► DIMM's at Slot 1, 2, 3 & 4 on the same memory card must be identical.		

2.3.3 Memory bandwidth

The POWER7 processor has exceptional cache, memory, and interconnect bandwidths. Table 2-6 shows the bandwidth estimates for the Power 710 and Power 730 systems.

Table 2-6 Power 710 and Power 730 memory and I/O bandwidth estimates

Memory	Bandwidth at processor core frequencies		
	3.0 GHz	3.55 GHz	3.7 GHz
L1 (data) cache	144 GB/s	170.4 GB/s	178.56 GB/s

Memory	Bandwidth at processor core frequencies		
	3.0 GHz	3.55 GHz	3.7 GHz
L2 cache	144 GB/s	170.4 GB/s	178.56 GB/s
L3 cache	96 GB/s	113.6 GB/s	119.04 GB/s
System memory	68.22 GB/s per socket	68.22 GB/s per socket	68.22 GB/s per socket

2.4 Capacity on Demand

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

2.5 Factory deconfiguration of processor cores

The Power 710 and Power 730 servers have the capability to be shipped with part of the installed cores deconfigured at the factory. The primary use for this feature is to assist with optimization of software licensing. A deconfigured core is unavailable for use in the system and thus does not require software licensing.

Feature #2319 deconfigures one core in a Power 710 or Power 730 system. It is available on all systems configurations.

The maximum number of this feature that can be ordered is one less than the number of cores in the system. For example a maximum of 5 x #2319 can be ordered for a 6-core system. Feature #2319 can only be specified at initial order and cannot be applied to an installed machine.

2.6 Technical comparison of Power 710 and Power 730

Table 2-7 shows a comparison of the technical aspects of the two systems, the Power 710 and the Power 730.

Table 2-7 Comparison of technical characteristics between Power 710 and Power 730

System characteristics	Power 710	Power 730
Processor modules	4-core 3.0 GHz POWER7 6-core 3.7 GHz POWER7 8-core 3.55 GHz POWER7	4-core 3.0 GHz POWER7 4-core 3.7 GHz POWER7 6-core 3.7 GHz POWER7 8-core 3.55 GHz POWER7
Pluggable processor modules	1 only	2
Min./Max. processor cores	4/8	8/8 (4-core modules) 12/12 (6-core modules) 16/16 (8-core modules)
L3 cache	on-chip eDRAM	on-chip eDRAM

System characteristics	Power 710	Power 730
Max memory slots and type	4 slots per memory riser card, (8 slots max.), DDR3 at 1066 MHz	4 slots per memory riser card, (16 slots max.), DDR3 at 1066 MHz
Memory chipkill	Yes	Yes
Memory spare	No	No
Memory hotplug	No	No
TPMD card	Yes	Yes
PCIe x8 slots (in chassis)	4 low profile/short cards	4 low profile/short cards
PCI-X 2.0 slots in chassis	n/a	n/a
PCIe and PCI-X hot plug	No	No
PowerVM support	Optional	Optional
Capacity on Demand	No	No
Redundant hotplug power	Optional	Yes
DASD bays (hot-plug, front access, SFF)	3 or 6	3 or 6
GX++ slots	1 for clustering support only	2 for clustering support only

2.7 System bus

This section provides additional information related to the internal buses.

The Power 710 and 730 systems have internal I/O connectivity through PCIe slots, as well as external connectivity through Infiniband adapters.

The internal I/O subsystem on the Power 710 and 730 is connected to the GX bus on a POWER7 processor in the system. This bus runs at 1.25 GHz and provides 10 GB/s of I/O connectivity to the PCIe slots, IVE ports, SAS internal adapters, and USB ports.

Additionally, the POWER7 processor chip 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 a 4x Infiniband adapter. Each bus runs at 2.5 GHz and provides 20 GB/s bandwidth.

The GX++ slots do not support hot plug. On the Power 710 and 730 systems, I/O drawers are not supported. The Dual-port 4x Channel Attach card (#5266) can be connected on the GX++ slots and be used to connect to High Performance Computing clusters, or commercial clusters such as the IBM DB2® pureScale™ solution.

Table 2-8 provides I/O bandwidth of the Power 710 and Power 730 systems.

Table 2-8 I/O bandwidth

I/O	Bandwidth
GX++ Bus	20 GB/sec (one on the Power 710, two on the Power 730)
GX+ Bus	10 GB/sec (Internal I/O)

I/O	Bandwidth
Total I/O	30 GB/sec for Power 710 50 GB/sec for Power 730

2.8 Internal I/O subsystem

The internal I/O subsystem resides on the system planar, which supports PCIe slots. PCIe slots on the Power 710 and 730 are not 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.

2.8.1 Slot configuration

Table 2-9 displays the slot configuration of Power 710 and Power 730.

Table 2-9 Slot configuration of a Power 710 and Power 730.

Slot#	Description	Location code	PCI Host Bridge (PHB)	Max. card size
Slot 1	PCIe x8	P1-C3	PCIe PHB0	Low Profile
Slot 2	PCIe x8	P1-C4	PCIe PHB1	Low Profile
Slot 3	PCIe x8	P1-C5	PCIe PHB2	Low Profile
Slot 4	PCIe x8	P1-C6	PCIe PHB3	Low Profile

2.8.2 System ports

The system planar has two serial ports that are called system ports. When a Hardware Management Console (HMC) is connected to the system, the integrated system ports are rendered nonfunctional. In this case, you must install an asynchronous adapter, which is described in Table 2-17 on page 47 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).

Note: 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.

2.9 Integrated Virtual Ethernet adapter

The POWER7 processor-based servers extend the virtualization technologies introduced in POWER5 by offering the Integrated Virtual Ethernet adapter (IVE). IVE, also sometimes known as Host Ethernet Adapter (HEA), enables an easy way to manage the sharing of the integrated high-speed Ethernet adapter ports. There are various Integrated Virtual Ethernet adapter card options:

- #1832 Quad-port 1 Gb HEA Daughter Card (Copper)
- #1833 Dual-port 10 Gb HEA Daughter Card (Fiber)
- #1837 Dual-port 10 Gb HEA Daughter Card (Copper)

The IVE includes special hardware features to provide logical Ethernet ports that can communicate to logical partitions (LPAR) reducing the use of IBM POWER Hypervisor™ (PHYP). Its design provides a direct connection for multiple LPARs, allowing LPARs to access external networks through the IVE without having to go through an Ethernet bridge on another logical partition, such as the Virtual I/O Server. Therefore, this eliminates the need to move packets (using Virtual Ethernet Adapters) between partitions and then through a Shared Ethernet Adapter (SEA) to an physical Ethernet port. LPARs can share IVE ports with improved performance.

2.9.1 IVE and SEA implementations

Figure 2-9 shows the difference between IVE and SEA implementations.

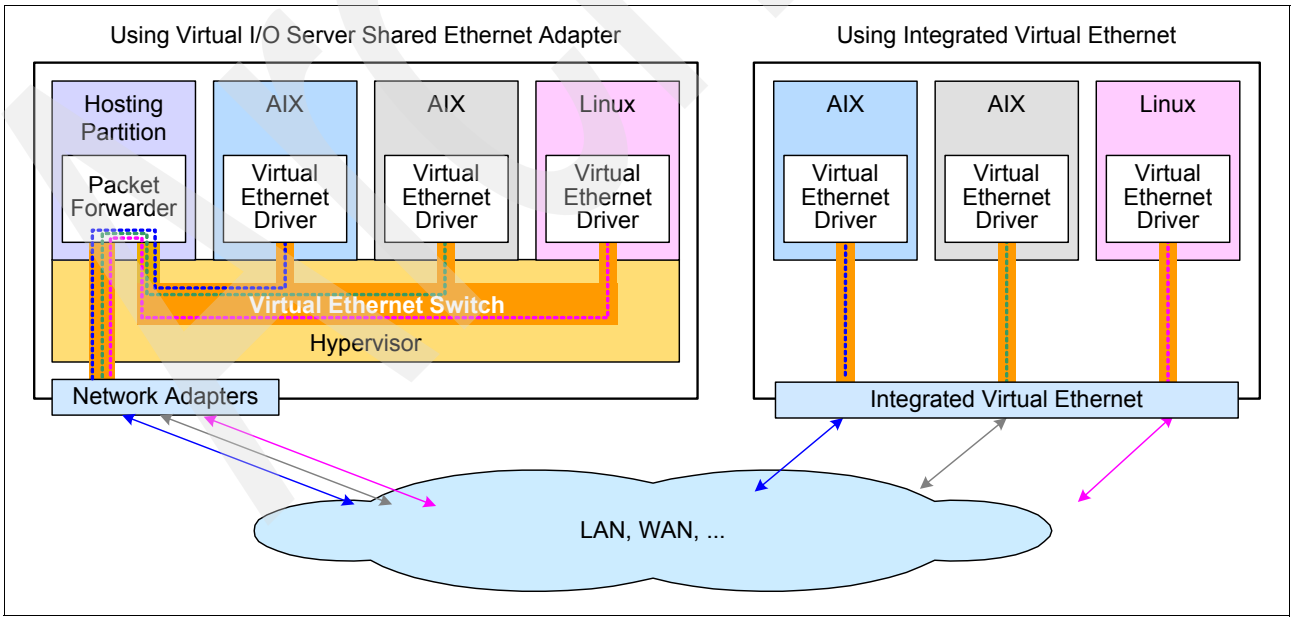


Figure 2-9 Integrated Virtual Ethernet compared to Virtual I/O Server Shared Ethernet Adapter

The IVE design provides high performance Ethernet with efficient virtualization. It offers:

- Either four 1 Gbps (#1832) or two 10 Gbps ports (#1833 - Fiber or #1837 - Copper)
- External network connectivity for LPARs using dedicated ports without the need of a Virtual I/O Server
- Industry standard hardware acceleration, with flexible configuration possibilities
- The speed and performance of the GX+ bus
- Great improvement of latency for short packets that are ideal for messaging applications such as distributed databases that require low latency communication for synchronization and short transactions

Table 2-10 lists the various IVE options available on the Power 710 and 730 systems.

Table 2-10 IVE options on the Power 710 and 730

Feature code	Description
#1832	Quad-port 1 Gb HEA Daughter Card
#1833	Dual-port 10 Gb HEA Daughter Card (Fiber)
#1837	Dual-port 10 Gb HEA Daughter Card (Copper)

For more information about IVE features, read *Integrated Virtual Ethernet Adapter Technical Overview and Introduction*, REDP-4340.

2.9.2 IVE subsystem

One of the key design goals of the IVE is the capability to integrate up to two 10 Gbps Ethernet ports or four 1 Gbps Ethernet ports into the P5IOC2 chip, with the effect of a low cost Ethernet solution for low-end and mid-range server platforms.

Each IVE can provide 32 logical ports to client partitions on the server. The quad-port 1 Gb adapter supports up to eight logical ports per physical port. The two dual-port 10 Gb adapters support up to 16 logical ports per physical port. If the IVE card is replaced, then a new IVE card provides a new set of MAC addresses.

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

2.10 PCI adapters

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 using two pairs of wires, one pair for transmit and one pair for receive, and can only transmit one bit per cycle. It can transmit at the extremely high speed of 2.5 Gbps, which equates to a burst mode of 320 MBps on a single connection. These two pairs of wires are called a lane. A PCIe link can be composed of multiple lanes. In such configurations, the connection is labeled as x1, x2, x8, x12, x16 or x32, where the number is effectively the number of lanes.

IBM offers only PCIe low profile adapter options for the Power 710 and 730 systems. All adapters support Extended Error Handling (EEH). PCIe adapters use another type of slot

than the PCI and PCI-X adapters. If you attempt to force an adapter into the wrong type of slot, you might damage the adapter or the slot.

Note: IBM i IOPs are not supported.

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 PTF 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 under various classifications (LAN, SCSI, SAS, Fibre Channel (FC), and so forth) and provide tables of orderable feature numbers. The tables indicate that the feature is supported by the AIX (A), IBM i (i), and Linux (L) operating systems.

2.10.1 LAN adapters

To connect to a local area network (LAN), you can use Integrated Virtual Ethernet. Other LAN adapters are supported in the system enclosure PCIe slots. Table 2-11 lists the additional LAN adapters that are available.

Table 2-11 Available LAN adapters

Feature code	Adapter description	Slot	Size	OS support
#5271	PCIe LP 4-Port 10/100/1000 Base-TX Ethernet Adapter	PCIe	Low profile Short	A, L
#5272	PCIe LP 10GbE CX4 1-port Adapter	PCIe	Low profile Short	A, L
#5274	PCIe LP 2-Port 1GbE SX Adapter	PCIe	Low profile Short	A, i, L
#5275	PCIe LP 10GbE SR 1-port Adapter	PCIe	Low profile Short	A, L

2.10.2 Graphics accelerators

Table 2-12 lists the available graphics accelerator. It can be configured to operate in either 8-bit or 24-bit color modes. The adapter supports both analog and digital monitors.

Table 2-12 Available graphics accelerators

Feature code	Adapter description	Slot	Size	OS support
#5269	PCIe LP POWER GXT145 Graphics Accelerator	PCIe	Low profile Short	A, L

2.10.3 SAS adapters

Table 2-13 lists the SAS adapter available for the Power 710 and Power 730 systems.

Table 2-13 Available SAS adapters

Feature code	Adapter description	Slot	Size	OS support
#5278	PCIe LP 2-x4-port SAS Adapter 3 Gb	PCIe	Low profile Short	A, i, L

2.10.4 PCIe RAID & SSD SAS adapter

A new SSD option for selected POWER7 processor-based servers offers a significant price/performance improvement for many client SSD configurations. The new SSD option is packaged somewhat differently from the currently available Power Systems SSDs. The new PCIe RAID & SSD SAS adapter has up to four 177 GB SSD modules plugged directly onto the adapter, thus saving the need for the SAS bays and cabling associated with the current SSD offering. Depending on the configurations being compared to the current 69 GB SAS-bay-based SSD offering, the new PCIe-based SSD offering can save up to 70% of the list price, assuming equivalent GB, and up to 65% smaller footprint, assuming equivalent GB.

Figure 2-10 shows the double-wide adapter and SSD modules.

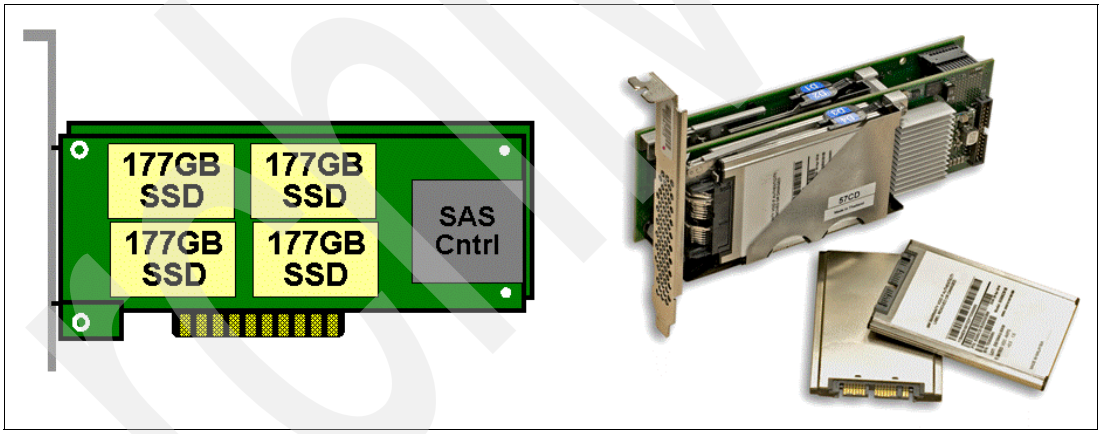


Figure 2-10 The PCIe RAID & SSD SAS adapter and 177 GB SSD modules

Table 2-14 shows the available RAID & SSD SAS adapter for the Power 710 and Power 730:

Table 2-14 PCIe RAID & SSD SAS adapter characteristics

Feature code	Adapter description	Slot	Size	OS support
#2053	PCIe LP RAID & SSD SAS Adapter 3 Gb	PCIe	Low profile Double wide, short	A, i, L

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

Because the SSD modules are mounted on the adapter, in order 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 removes RAID error correcting information, so it is best to mirror the data using operating system tools in order to prevent data loss in case of failure.

2.10.5 Fibre Channel adapters

The systems support direct or SAN connection to devices using Fibre Channel adapters. Table 2-15 provides a summary of the available Fibre Channel adapters.

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

Table 2-15 Available Fibre Channel adapters

Feature code	Adapter description	Slot	Size	OS support
#5273	PCIe LP 8 Gb 2-Port Fibre Channel Adapter	PCIe	Low profile Short	A, i, L
#5276	PCIe LP 4 Gb 2-Port Fibre Channel Adapter	PCIe	Low profile Short	A, i, L

Note: NPIV through the Virtual I/O server is only supported with adapter #5273.

2.10.6 Fibre Channel over Ethernet

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

Figure 2-11 shows a comparison between existing FC and network connections and FCoE connections.

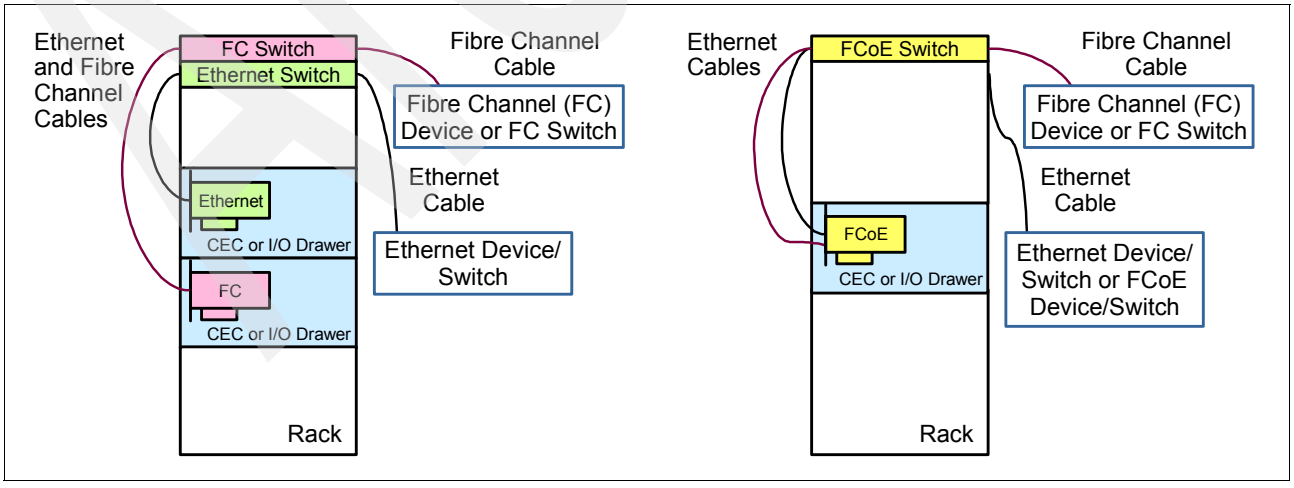


Figure 2-11 Comparison between existing FC and network connection and FCoE connection

Table 2-16 lists the available Fibre Channel over Ethernet Adapter. It is a high performance, Converged Network Adapter (CNA) using SR optics. Each port can provide Network Interface Card (NIC) traffic and Fibre Channel functions simultaneously.

Table 2-16 Available FCoE adapter.

Feature code	Adapter description	Slot	Size	OS support
#5270	PCIe LP 10Gb FCoE 2-port Adapter	PCIe	Low profile Short	A, L

2.10.7 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 has been 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 bi-directional 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 as well as 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, read *HPC Clusters Using InfiniBand on IBM Power Systems Servers*, SG24-7767.

IBM offers the GX++ Dual-port 4x Channel Attach (#5266) that plugs into a GX++ slot in the system enclosure. The adapter provides two 4X connections for 4X channel applications. This adapter is only used for server clustering configurations. One GX++ slot is available on the Power 710. Two GX++ slots are available on the Power 730. Feature #5266 can be installed in either GX++ slot. AIX, IBM i, and Linux operating systems are supported.

2.10.8 Asynchronous adapters

Asynchronous PCI 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 feature listed in Table 2-17.

Table 2-17 Available asynchronous adapters

Feature code	Adapter description	Slot	Size	OS support
#5277	PCIe LP 4-Port Async EIA-232 Adapter	PCIe	Low profile Short	A, L

2.11 Internal storage

The Power 710 and Power 730 servers use an integrated SAS/SATA connected to a 66 MHz PCI-X bus on the P5-IOC2 chip (see Figure 2-12). The SAS/SATA controller used in the server's enclosure has two sets of four SAS/SATA channels. 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:

- Feature #5263 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 RAID.

  Note: The feature #5623 is not supported with IBM i.

- Feature #5267 supports six SFF disk units, either HDD or SSD, and a SATA DVD. There is no support for split backplane and RAID.
- Feature #5268 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. There is no support for split backplane and RAID. This feature is required when IBM i is the primary operating system (#2145).

Note: One of features #1883, #1884, #1885, #1886, #1888, #1890, #1909, or #1911 must be selected (no HDDs/SSDs are required in the CEC if feature #0837 is selected).

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 2-12 details the internal topology overview for the #5263 backplane:

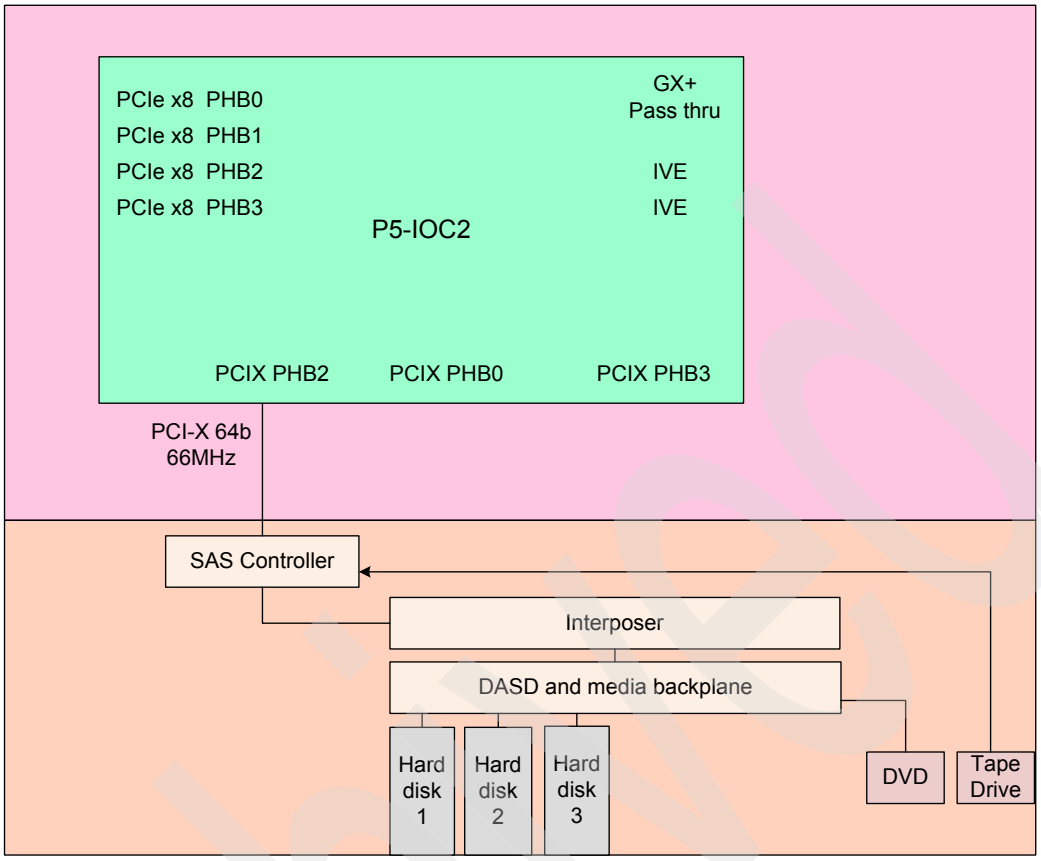


Figure 2-12 Internal topology overview for #5263 DASD backplane

Figure 2-13 shows the internal topology overview for the #5267 backplane:

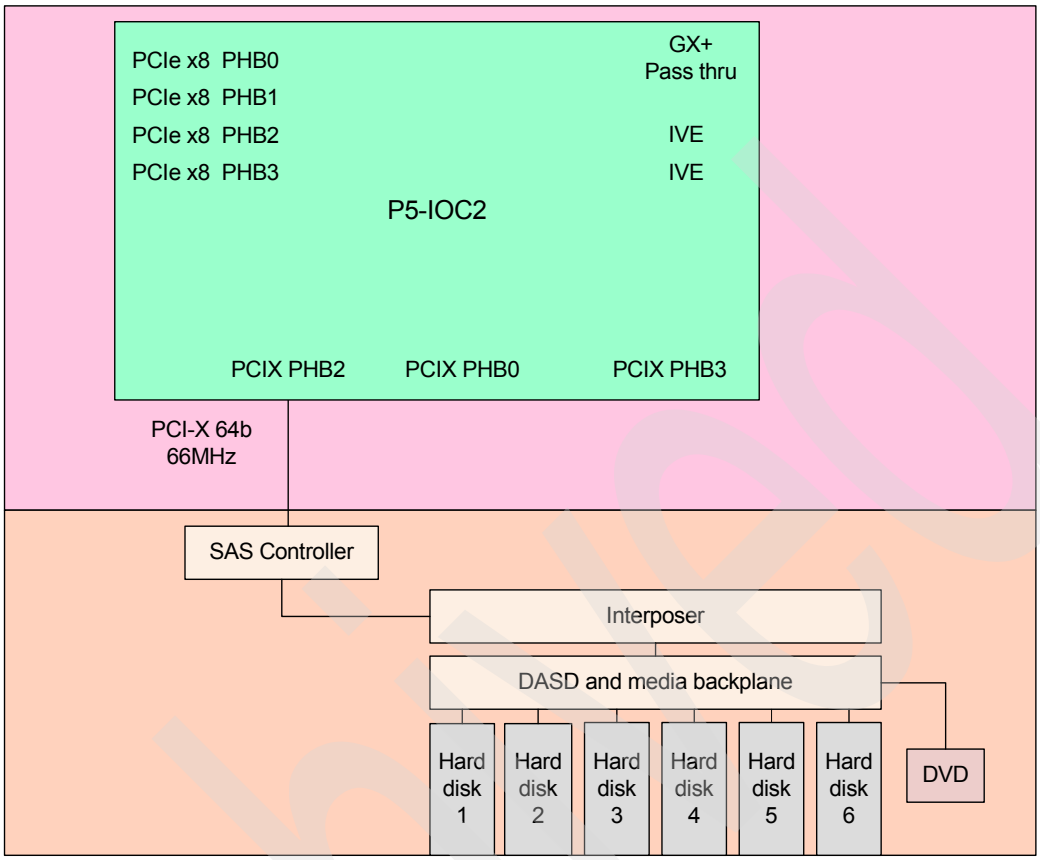


Figure 2-13 Internal topology overview for the #5267 DASD backplane

Figure 2-14 explains the details of the internal topology overview for the #5268 backplane:

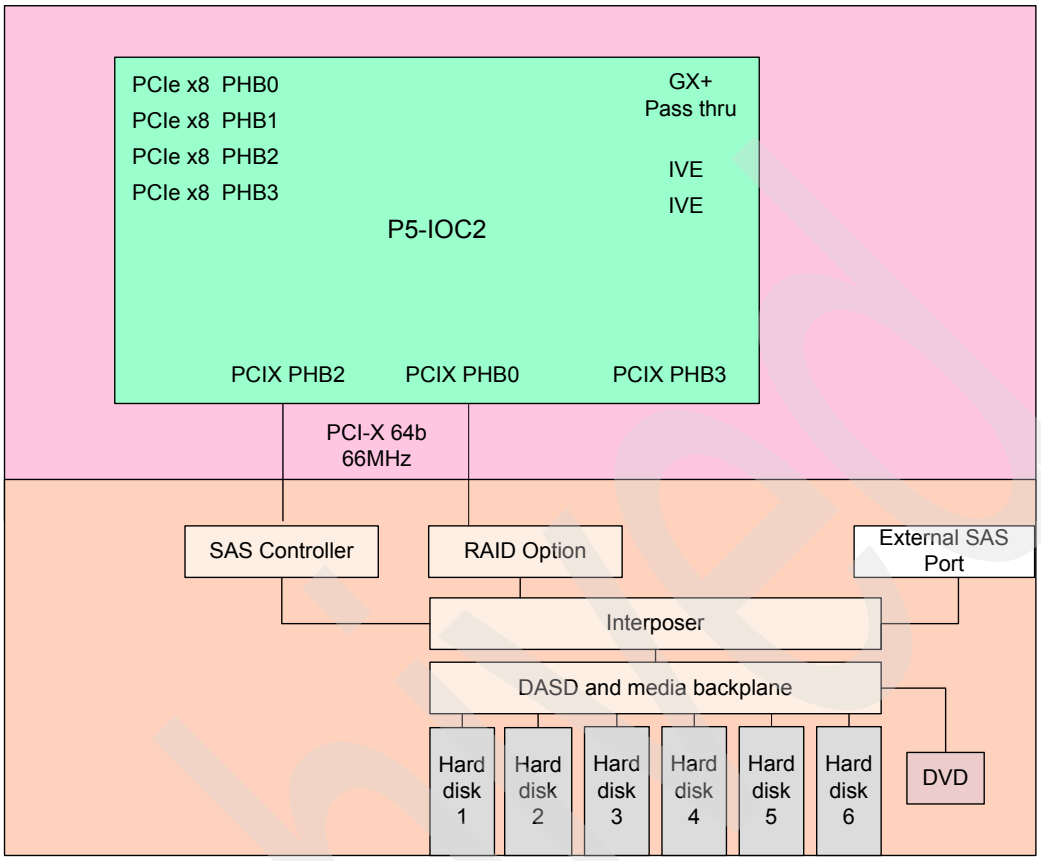


Figure 2-14 Internal topology overview for the #5268 DASD backplane

2.11.1 RAID support

There are multiple protection options for HDD/SSD drives in the SAS SFF bays in the Power 710 and 730 system unit or drives in disk drawers or drives in disk-only I/O drawers. Although protecting drives is always the best idea, 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. So IBM i also supports controllers providing 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 provided in the Power 710 and Power 730 system units. It is optionally augmented by RAID 5 and RAID 6 capability when #5268 is added to the configuration. Without #5268, the integrated controller supports system mirroring protection for AIX, IBM i, and Linux and supports RAID 0 or 10 protection for AIX and Linux. Other disk and SSD controllers are available as PCIe SAS adapters. PCI controllers with and without write cache are supported, as well as RAID 5 and RAID 6 on controllers with write cache.

AIX and Linux can use disk drives 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 around 283 GB. IBM i always uses drives formatted to 528 bytes. IBM Power SSDs are formatted to 528 bytes.

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

Supported RAID functions

The RAID enablement feature and the auxiliary write cache are optional features. When features are configured, Power 710 and 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 will result in the loss of all data on the array. This increases I/O bandwidth by simultaneously accessing multiple data paths.

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

  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.

- 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, which makes it less cost effective than RAID-5 for equivalent storage capacity.

- RAID-10 is also known as a stripe 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 2-disk array for performance benefits. A duplicate of the first stripe set is then mirrored on another 2-disk array for fault tolerance.

2.11.2 External SAS port

The Power 710 and Power 730 DASD backplane (#5268) offers the 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, the System Storage 7214 Tape and DVD Enclosure (Model 1U2), and the IBM System Storage 7216 Multi-Media Enclosure (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.

2.11.3 Media bays

The Power 710 and 730 each have a slim media bay that contains an optional DVD-RAM (#5762) and a tape bay (only available with #5263) that can contain a tape drive or removable disk drive. Direct dock and hot-plug of the DVD media device is supported.

2.12 External I/O subsystems

The Power 710 and the Power 730 systems do not support attachment of I/O drawers.

2.13 External disk subsystems

The following external disk subsystems can be attached to the Power 710 and Power 730 server:

- EXP 12S SAS Expansion Drawer (#5886 - Supported, but is no longer orderable)
- EXP24S SFF Gen2-bay Drawer (#5887)
- IBM System Storage

The next sections describe the EXP 12S AND EXP24S Expansion Drawers and IBM System Storage in more detail.

Note: The EXP 12S SAS Expansion Drawer (#5886) is not supported on a 4-core Power 710.

2.13.1 EXP 12S SAS Expansion Drawer

The EXP 12S (#5886) is an expansion drawer with twelve 3.5-inch form factor SAS bays. The #5886 supports up to 12 hot-swap SAS Hard Disk Drives (HDD) or up to 8 hot-swap Solid State Drives (SSD). The EXP 12S includes redundant ac power supplies and two power cords. Though 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 EXP 12S takes up a 2 EIA space in a 19-inch rack. The SAS controller can be a SAS PCI-X or PCIe adapter or pair of adapters.

The drawer can either be attached using the #5268 storage backplane, providing an external SAS port, or using a LP 2-x4-port SAS adapter 3 Gb (#5278).

With proper cabling and configuration, multiple wide ports are used to provide redundant paths to each dual port SAS disk. The adapter manages SAS path redundancy and path switching in case a SAS drive failure occurs. The SAS Y cables attach to an EXP 12S Disk Drawer. Use SAS cable (YI) system to SAS enclosure, single controller/dual path 1.5M (#3686 - supported but not longer orderable) or SAS cable (YI) system to SAS enclosure, single controller/dual path 3M (#3687) to attach SFF SAS drives in an EXP12S Drawer.

Figure 2-15 on page 53 illustrates connecting a system external SAS port to a disk expansion drawer

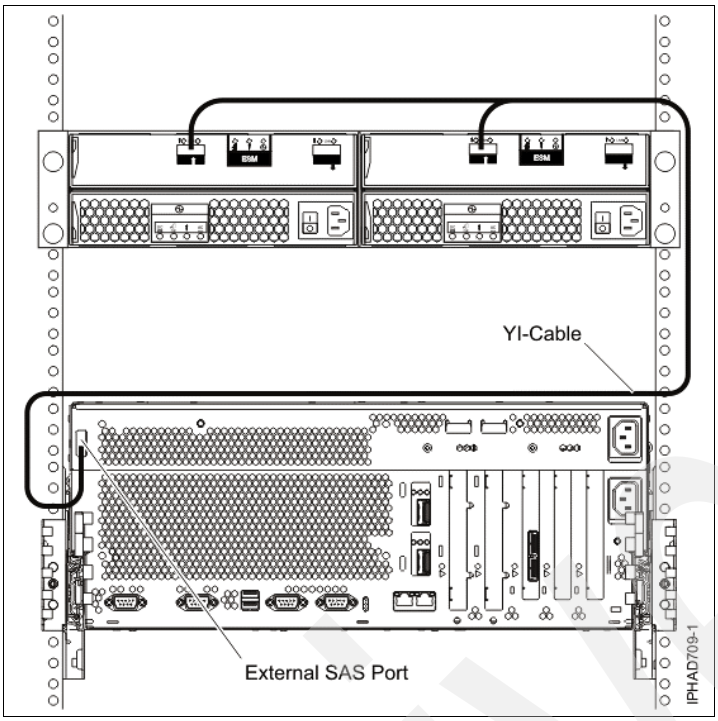


Figure 2-15 External SAS Cabling

Use SAS cable (YO) system to SAS enclosure, single controller/dual path 1.5M (#3450) or SAS cable (YO) system to SAS enclosure, single controller/dual path 3M (#3451) to attach SFF SAS drives in an EXP12S Drawer.

In the EXP 12S Drawer, a high availability I/O configuration can be created using a pair of #5278 adapters and SAS X cables to protect against the failure of a SAS adapter.

A second EXP12S drawer can be attached to another drawer using two SAS EE cables, providing 24 SAS bays instead of 12 bays for the same SAS controller port. This arrangement is called cascading. In this configuration, all 24 SAS bays are controlled by a single controller or a single pair of controllers.

The EXP12S Drawer can also be directly attached to the SAS port on the rear of the Power 710 or Power 730, providing a very low cost disk storage solution. The rear SAS port is provided by the Storage Backplane - 6 SFF Drives/SATA DVD/RAID/External SAS Port (#5268). A second unit cannot be cascaded to a EXP12S Drawer attached in this way.

Note: If the internal disk bay of the Power 710 or Power 730 contains any SSD drives, an EXP 12S 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 only contains SAS disk drives).

For detailed information about the SAS cabling, see the serial-attached SCSI cable planning documentation at:

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

2.13.2 EXP24S SAS Expansion Drawer

The EXP24S (#5887) SFF Gen2-bay Drawer is an expansion drawer with twenty-four 2.5-inch small form factor (SFF) SAS bays. It supports up to 24 hot-swap SFF SAS Hard Disk drives (HDD) on POWER6 or POWER7 servers in 2U of 19-inch rack space.

Note: SSD not currently supported

The SFF bays of the EXP24S are different from the SFF bays of the POWER7 system units or 12X PCIe I/O Drawers (#5802, #5803). The EXP24S uses Gen-2 or SFF-2 SAS drives that physically do not fit in the Gen-1 or SFF-1 bays of the POWER7 system unit or 12X PCIe I/O Drawers or viceversa.

The EXP24S SAS ports are attached to SAS controllers which can be a SAS PCI-X or PCIe adapter or pair of adapters. The EXP24S can also be attached to an imbedded SAS controller in a server with an imbedded SAS port. Attachment between the SAS controller and the EXP24S SAS ports is via the appropriate SAS Y or X cables.

The drawer can either be attached using the #5268 storage backplane, providing an external SAS port, or using a LP 2-x4-port SAS adapter 3 Gb (#5278).

Note: A single #5887 drawer can be cabled to the CEC external SAS port when a #5268 DASD backplane is part of the system. A 3Gb/s YI cable (#3686/#3687) is used to connect a #5887 to the CEC external SAS port.

A single #5887 will not be allowed to attach to the CEC external SAS port when a #8350 processor (4-core) is ordered/installed on a single socket Power 710 system.

The EXP24S can be ordered in one of 3 possible manufacturing-configured MODE settings (not customer set-up).

- 1, 2 or 4 sets of disk bays

With IBM AIX/Linux/VIOS, the EXP24S can be ordered with four sets of 6 bays (mode4), two sets of 12 bays (mode 2) or one set of 24 bays (mode 1). With IBM i the EXP24S can be ordered as one set of 24 bays (mode 1).

Note:

- #5887 Modes are set by IBM Manufacturing. No option to reset after IBM ships is announced.
- If ordering multiple EXP24S avoid mixing modes within that order. There is no externally visible indicator as to the drawer's mode.
- SSD are not currently supported in EXP24S .
- No Cascading of #5887s - Max one #5887 per SAS connector
- Power 710 or Power 730 can support up to four #5887

There are six SAS connectors on the rear of the EXP24S to which SAS adapters/controllers are attached. They are labeled T1, T2, and T3, and there are two T1, two T2, and two T3, details on Figure 2-16.

- 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 in order to access allSAS bays.

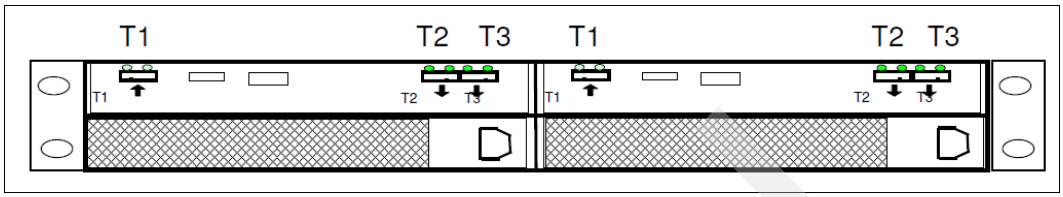


Figure 2-16 #5887 rear connectors

An EXP24S in mode 4 can be attached to two or four SAS controllers and provide a great deal 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/controllers.

EXP24S no-charge specify codes should be included with EXP24S orders to indicate IBM Manufacturing the mode to which the drawer should be set and the adapter/controller/cable configuration which will be used. Table 2-18 lists the no-charge specify codes, the physical adapters/controllers/cables with their own chargeable feature numbers.

Table 2-18 EXP24S Cabling

Feature code	Mode	Adapter/controller	Cable to Drawer	Environment
#9359	1	One #5278	1 YO Cable	A, L,VIOS
#9360	1	Pair #5278	2 YO Cables	A, L,VIOS
#9361	2	Two #5278	2 YO Cables	A, L,VIOS
#9365	4	Four #5278	2 X Cable	A, L,VIOS
#9366	2	Two pair #5278	2 X Cables	A, L,VIOS
#9384	1	CEC SAS port	1 YI Cable	A,i, L,VIOS

Cable to EXP24S Drawer options:

- X cables for #5278 : 3 m (#3661), 6 m (#3662), 15 m (#3663)
- YO cables for #5278: 1.5 m (#3691), 3 m (#3692), 6 m (#3693), 15 m (#3694)
- YI cables for system unit SAS port (3 Gb): 1.5 m (#3686) or 3 m (#3687)

Note: EXP24S Drawer rails are fixed lenght and designed to fit Power Systems provided racks of 28 inches (8xx mm) deep. EXP24S Uses 2 EIA of space in a 19-inch wide rack. Other racks may have different depths and these rails won't adjust. No adjustable depth rails are orderable at this time

For detailed information about the SAS cabling, see the serial-attached SCSI cable planning documentation at:

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

2.13.3 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 up to high-end storage systems.

IBM System Storage N series

The IBM System Storage N series is a Network Attached Storage (NAS) solution and provides the latest technology to customers to help them 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:

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

IBM System Storage DS3000 family

The IBM System Storage DS3000 is an entry-level storage system designed to meet the availability and consolidation needs for a wide range of users. New features, including larger capacity 450 GB SAS drives, increased data protection features such as RAID 6, and more FlashCopies per volume, provide a reliable virtualization platform. For more information about the DS3000 family, see:

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

IBM System Storage DS5000

New DS5000 enhancements help reduce cost by introducing SSD drives. Also with the new EXP5060 expansion unit supporting 60 1 TB SATA drives in a 4U package, customers can see up to a 1/3 reduction in floor space over standard enclosures. With the addition of 1 Gbps iSCSI host attach, customers can reduce cost for their less demanding applications while continuing to provide high performance where necessary, utilizing the 8 Gbps FC host ports. With the DS5000 family, you get consistent performance from a smarter design that simplifies your infrastructure, improves your TCO, and reduces your cost. For more information about the DS5000 family, see:

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

IBM Storwize V7000 Midrange Disk System

IBM® Storwize® V7000 is a virtualized storage system to complement virtualized server environments that provides unmatched performance, availability, advanced functions, and highly scalable capacity never seen before in midrange disk systems. Storwize V7000 is a powerful midrange disk system that has been designed to be easy to use and enable rapid deployment without additional resources. Storwize V7000 is virtual storage that offers greater efficiency and flexibility through built-in solid state drive (SSD) optimization and thin provisioning technologies. Storwize V7000 advanced functions also enable non-disruptive migration of data from existing storage, simplifying implementation and minimizing disruption to users. Storwize V7000 also enables you to virtualize and reuse existing disk systems, supporting a greater potential return on investment (ROI). For more information about Storwize V7000, see:

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

IBM XIV Storage System

IBM offers a mid–sized configuration of its self-optimizing, self-healing, resilient disk solution, the IBM XIV® Storage System: storage reinvented for a new era. Now, organizations with mid-size capacity requirements can take advantage of the IBM latest technology for their

most demanding applications with as little as 27 TB usable capacity and incremental upgrades. For more information about XIV, see:

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

IBM System Storage DS8000

The IBM System Storage DS8000® family is designed to offer high availability, multiplatform support, and simplified management tools. With its high capacity, scalability, broad server support, and virtualization features, the DS8000 family is well suited for simplifying the storage environment by consolidating data from multiple storage systems on a single system.

The high-end model DS8700 is the most advanced model in the IBM DS8000 family lineup and introduces new dual IBM POWER6-based controllers that usher in a new level of performance for the company's flagship enterprise disk platform. The DS8700 offers twice the maximum physical storage capacity than the previous model. For more information about the DS8000 family, see:

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

2.14 Hardware Management Console

The Hardware Management Console (HMC) is a dedicated workstation that provides a graphical user interface for configuring, operating, and performing basic system tasks for the POWER7 processor-based systems, as well as the previous POWER5, POWER5+, POWER6, and POWER6+ processor-based systems. Either non-partitioned, partitioned, or clustered environments are supported. In addition, the HMC is used to configure and manage logical partitions (LPARs). One HMC is capable of controlling multiple POWER5, POWER5+, POWER6 and POWER6+ and POWER7 processor-based systems.

At the time of writing, one HMC supports up to 1024 LPARs using HMC machine code Version 7 Release 710 or later. It can also support up to 48 Power 710, Power 720, Power 730, Power 740, Power 750, Power 755, Power 770, or Power 780. Up to 32 Power 795 servers are supported. For updates of the machine code and HMC functions and hardware prerequisites, see the following website:

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

2.14.1 HMC functional overview

The HMC provides three groups of tasks, as described in the following topics.

Server management

The first group contains all tasks related to the management of the physical servers under the control of the HMC:

- System password
- Power on/off
- Capacity on Demand
- Error management
  - System indicators
  - Error/event collection reporting
  - Dump collection reporting
  - Call home
  - Customer notification

  - Hardware replacement (guided repair)
  - SNMP events
- Concurrent add/repair
- Redundant Service Processor
- Firmware updates

Virtualization management

The second group contains all tasks related to virtualization features such as the Logical Partition configuration or dynamic reconfiguration of resources:

- System Plans
- System Profiles
- Logical Partitions (create, activate, shutdown)
- Profiles
- Partition Mobility
- Dynamic LPAR operations (add/remove/move processors, memory, I/O, and so on.)
- Custom Groups

HMC console management

The last group relates to the management of the HMC itself. This includes its maintenance, security, or configuration. Here are a few examples:

- Set-up wizard
- User management:
  - User IDs
  - Authorization levels
  - Customizable authorization
- Disconnect/reconnect
- Network security:
  - Remote operation enable/disable
  - User definable SSL certificates
- Console logging
- HMC redundancy
- Scheduled operations
- Back up and restore
- Updates, upgrades
- Customizable message of the day

The HMC version V7R720 adds support for Power 710, Power 720, Power 730, Power 740, and Power 795 servers.

The HMC provides both graphical and command line interface for all management tasks. Running HMC Version 7, a remote connection to the HMC using a web browser or SSH is possible. A command line interface is also available by using the SSH secure shell connection to the HMC. It can be used by an external management system or a partition to perform HMC operations remotely.

2.14.2 HMC connectivity to the POWER7 processor based systems

POWER5, POWER5+, POWER6, POWER6+, and POWER7 processor-based servers managed by an HMC require Ethernet connectivity between the HMC and the server 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. The rack-mounted 7042-CR5 and 7042-CR6 HMC default configuration provides four Ethernet ports. The deskside 7042-C07 and 7042-C08 HMC standard configuration offers one Ethernet port. Preferably, order an optional PCIe adapter to provide additional Ethernet ports.

For any logical partition in a server, it is possible to use a Shared Ethernet Adapter in the Virtual I/O Server or Logical Ports of the Integrated Virtual Ethernet card, for fewer connections from the HMC to partitions. Therefore, a partition does not require its own physical adapter to communicate to an HMC.

It is a good practice to connect the HMC to the first HMC port on the server, which is labeled as HMC Port 1, although other network configurations are possible. You can attach a second HMC to HMC Port 2 of the server for redundancy. Figure 2-17 shows a simple network configuration to enable the connection from an HMC to a server in order to enable Dynamic LPAR operations. For more details about HMC and the possible network connections, see the *Hardware Management Console V7 Handbook*, SG24-7491.

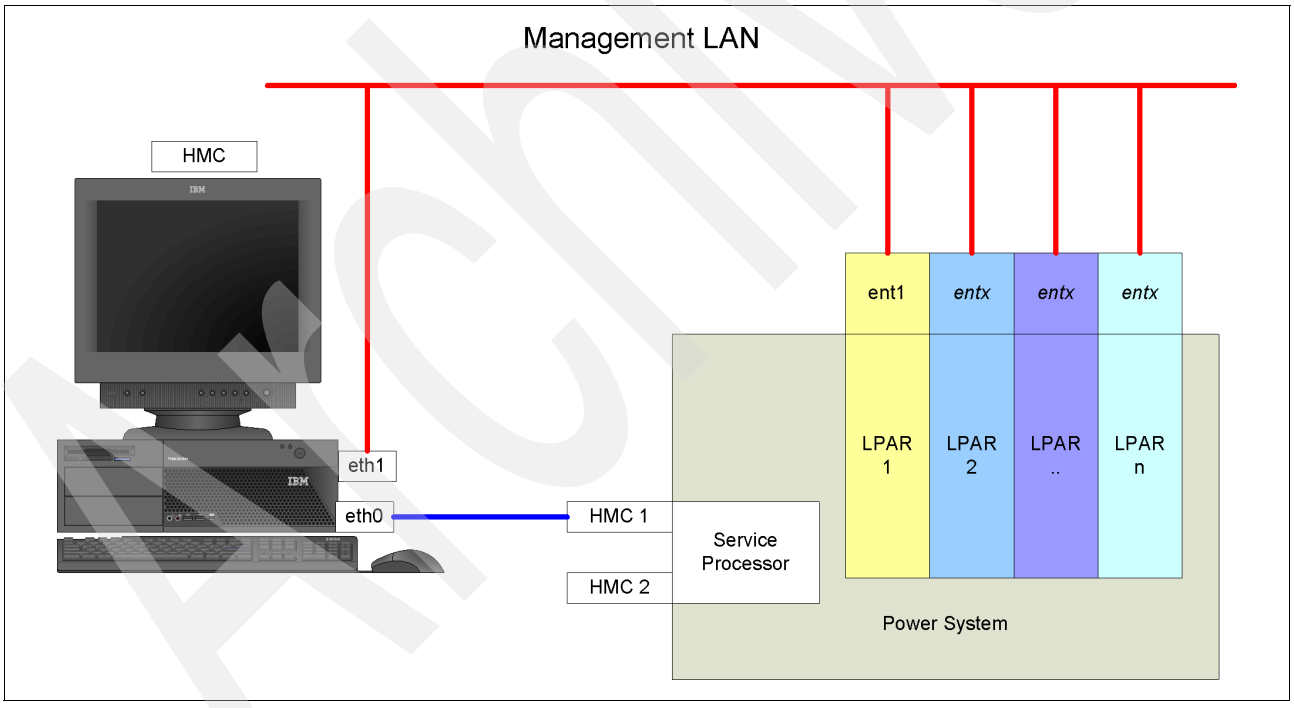


Figure 2-17 HMC to service processor and LPARs network connection

The default mechanism for allocation of the IP addresses for the service processor HMC ports is dynamic. The HMC can be configured as a DHCP server, providing the IP address at the time the managed server is powered on. In this case, the Flexible Service Processor (FSP) is allocated IP address from a set of address ranges predefined in the HMC software. These predefined ranges are identical for version 720 of the HMC code and for previous versions.

If the service processor of the managed server does not receive the DHCP reply before time-out, predefined IP addresses will be set up on both ports. Static IP address allocation is also an option. You can also configure the IP address of the service processor ports with a static IP address by using the Advanced System Management Interface (ASMI) menus.

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

Two Ethernet 10/100 Mbps ports are provided:

- Both Ethernet ports are only visible to the service processor and can be used to attach the server to an HMC or to access the Advanced System Management Interface (ASMI) options from a client web browser, using the http server integrated into the service processor internal operating system.
- When not configured otherwise (DHCP or from a previous ASMI setting), both Ethernet ports of the first FSP have predefined IP addresses:
  - The service processor Eth0 or HMC1 port is configured as 169.254.2.147 with netmask 255.255.255.0.
  - The service processor Eth1 or HMC2 port is configured as 169.254.3.147 with netmask 255.255.255.0.

More information about the Service Processor can be found in “Service processor” on page 118.

2.14.3 HMC high availability

The Hardware Management Console (HMC) is an important hardware component. When put into operation, POWER7 processor-based 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 DLPAR reconfiguration, a partition migration using PowerVM Live Partition Mobility, or the creation of a new partition. You can therefore choose to install two HMCs in a redundant configuration so that one HMC is always operational. For example, when performing maintenance of one HMC, the other is still operational.

If redundant HMC function is desired, the servers can be attached to two separate HMCs to address availability requirements. Both HMCs must have the same level of Hardware Management Console Licensed Machine Code Version 7 (#0962) to manage POWER7 processor-based servers or an environment with a mixture of POWER5, POWER5+, POWER6, POWER6+ 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. Depending on your environment, you have multiple options to configure the network.

Figure 2-18 shows one possible highly available HMC configuration managing two servers. Each HMC is connected to one FSP port of all managed servers.

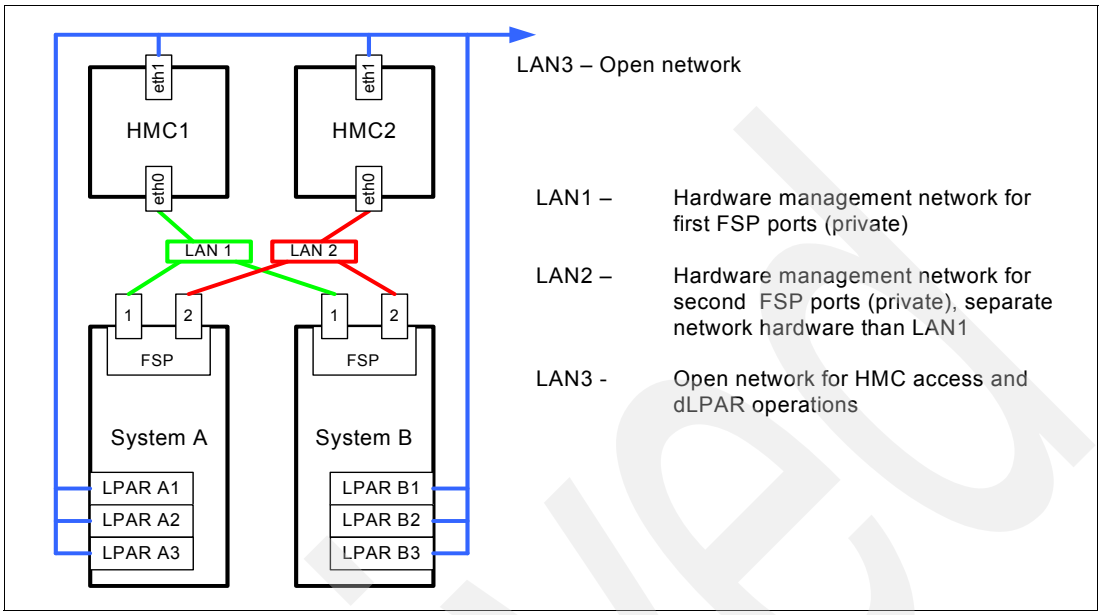


Figure 2-18 Highly available HMC and network architecture

Only the hardware management networks (LAN1 and LAN2) are shown as highly available in this diagram for simplicity. However, the management network (LAN3) can be made highly available by using a similar concept and adding more Ethernet adapters to LPARs and HMCs.

Both HMCs must be on a separate VLAN, to protect from a network failure. Each HMC can be a DHCP server for its VLAN.

In a configuration with multiple systems or HMCs, the customer is required to provide switches or hubs to connect each HMC to the server FSP Ethernet ports in each system. One HMC must connect to the port labeled as HMC Port 1, and a second HMC must be attached to HMC Port 2. This provides redundant connections for the HMCs and the service processor.

For more details about redundant HMCs, see the *Hardware Management Console V7 Handbook*, SG24-7491.

2.14.4 HMC code level

The HMC code must be at level V7R720 in order to support the Power 710 and Power 730 systems. In a dual HMC configuration, both HMCs must be at the same version and release of the HMC.

Tip: When upgrading the code of an HMC in a dual HMC configuration, it is a good practice to disconnect one of the HMCs and avoid having both HMCs connected to the same server with various code levels. If no profile or partition changes take place during the upgrade, both HMCs can stay connected. If the HMCs are at various code levels and a profile change is performed from the HMC at the higher code level, the format of the data stored in the server can change, and the HMC at the lower code level might go into recovery state if it does not understand the new data format.

Tip: There are compatibility rules between the various software products that are executing within a POWER7 processor-based server environment: HMC, Virtual I/O Server, system firmware, or partition operating systems. To check what combinations are supported, and to identify required upgrades, you can use the Fix Level Tool web page:

http://www14.software.ibm.com/webapp/set2/flrt/home

Here we list two important rules related to HMC code level when using PowerVM Live Partition Mobility:

- To use PowerVM Live Partition Mobility between a POWER6 processor-based server and a POWER7 processor-based server, if the source server is managed by one HMC and the destination server is managed by another HMC, then ensure that the HMC managing the POWER6 processor-based server is at version 7, release 3.5 or later, and the HMC managing the POWER7 processor-based server is at version 7, release 7.1 or later.
- To use PowerVM Live Partition Mobility for a partition configured for Active Memory Expansion, ensure that the HMC that manages the destination server is at version 7, release 7.1 or later.

2.15 Integrated Virtualization Manager

The Integrated Virtualization Manager (IVM) is a simplified hardware management solution that inherits most of the HMC features. The IVM runs within the Virtual I/O Server partition on the managed server. It manages a single server, avoiding the need of an independent appliance. It is designed to provide a solution that enables the administrator to reduce system setup time and to make hardware management easier, at a lower cost.

IVM provides a management model for a single system. Although it does not offer all of the HMC capabilities, it enables the exploitation of PowerVM technology. IVM targets the small and medium systems environment. The following POWER7 processor-based servers can be managed by IVM:

- Power 710
- Power 720
- Power 730
- Power 740
- Power 750

IVM is an addition to the Virtual I/O Server (VIOS), the product that enables I/O virtualization in the family of POWER processor-based systems. The IVM functions are provided by software executing within the Virtual I/O Server partition installed on the server to manage.

Note: In order to manage a Power 710 and Power 730 system through IVM, the Virtual I/O Server Version 2.2 is required.

Table 2-19 lists a detailed comparison between the IVM and the HMC functions.

Table 2-19 Comparison between IVM and HMC

Characteristics and functions	IVM	HMC
General characteristics		
Delivery Vehicle	Integrated into the server	A desktop or rack-mounted appliance
Installation	Installed with the Virtual I/O Server (optical or network). The preinstall option is available on certain systems.	Appliance is preinstalled. Reinstallation by optical media or network is supported.
Multiple system support	One IVM per server	One HMC can manage multiple servers
User Interface	Web browser (no local graphical display) and telnet session	Web browser (local or remote) or SSH access
Scripting and Automation	VIOS command line interface (CLI) and HMC compatible CLI	HMC command line interface
RAS characteristics		
Redundancy/HA of manager	One IVM per server	Multiple HMCs can manage the same system for HMC redundancy
Multiple VIOS	No – single VIOS	Yes
Fix/Update process for Manager	VIOS fixes and updates	HMC e-fixes and release updates
Adapter microcode updates	Inventory scout by RMC	Inventory scout by RMC
Firmware updates	Inband by OS – not concurrent	Service IBM Focal Point™ with concurrent firmware maintenance
I/O Concurrent Maintenance	VIOS support for slot and device level concurrent maintenance by the diag hot plug support	Guided support in the “Repair and Verify” function on the HMC
Serviceable event management	Service Focal Point Light - Consolidated management of firmware and management partition detected errors	Service Focal Point support for consolidated management of operating system and firmware detected errors
PowerVM function		
Full PowerVM Capability	Partial	Full
Capacity on Demand	Entry of PowerVM codes only	Full Support
I/O Support for IBM i	Virtual Only	Virtual and Direct
Multiple Shared Processor Pool	No - default pool only	Yes
Workload Management (WLM) Groups Supported	One	254

Characteristics and functions	IVM	HMC
Support for multiple profiles per partition	No	Yes
SysPlan Deploy & mksysplan	Yes	Yes

For a complete description of the possibilities offered by IVM, see *Integrated Virtualization Manager on IBM System p5*, REDP-4061.

2.15.1 IBM Systems Director Management Console (SDMC)

IBM Systems Director Management Console (SDMC) V6.7.3 is the next-generation Hardware Management Console (HMC). The SDMC incorporates the IBM Systems Director software to provide consistency with IBM Systems Director on other platforms.

The SDMC is available on physical hardware that is identical to that of the current HMC with additional memory and disk resources. SDMC is also available in a virtual appliance format for managing small-tier servers. The SDMC offers several advantages over the HMC, including the ability to manage both Power Systems servers and Power Systems blades based on Power Architecture®.

The SDMC is intended to be used in the same manner as the HMC. It provides the same functionality, including hardware, service, and virtualization management, for your Power Systems server and Power Systems blades. Because SDMC uses IBM Systems Director Express® Edition, it also provides all Systems Director Express capabilities, such as monitoring of operating systems and creating event action plans.

The SDMC can be obtained as a hardware appliance in the same manner as an HMC. Hardware appliances support managing Power Systems servers. The SDMC can optionally be obtained in a virtual appliance format, capable of running on VMware (ESX/i 4, or later), and KVM (Red Hat Enterprise Linux® (RHEL) 5.5). The virtual appliance is only supported for managing small-tier Power® servers and Power Systems blades. Much of the SDMC function is equivalent to the HMC. This includes:

- Server (host) management: Like HMC, the SDMC server (host) management functions include power control, inventory collection, firmware management including concurrent updates, error reporting, dump retrieval and transmission, error reporting via Call Home, and component repair and verify.
- Virtualization management: Like HMC, the SDMC virtualization management function includes management of logical partitions (virtual servers) and virtual I/ O, Micro-PartitioningTM, Partition (Virtual Server) Mobility, Partition (virtual server) Suspend and Resume, Active MemoryTM Sharing, N_Port ID Virtualization (NPIV), HCA (IB) configuration, and other PowerVMTM capabilities.
- Transition of configuration data: All partition (virtual server) related information is stored on the managed system itself and will be maintained as is. No configuration changes are required when a client moves from HMC management to SDMC management.
- Redundancy and high availability: The SDMC offers console redundancy similar to the HMC. The SDMC may be used in either an SDMC-SDMC pair or an SDMC- HMC pair. Alternatively, the SDMC may be configured as an Active-Passive High Availability pair. The Active-Passive HA function provides high availability for Systems Director operating system management functions.

While most functions that SDMC provides are very similar to that of the HMC, the SDMC has made significant strides in simplifying virtualization management. Specifically, managing virtual I/O has become much simpler as SDMC can automatically create virtual I/O adapters, and allow the user to assign virtual storage directly to partitions (virtual servers). SDMC provides a more intuitive interface for dynamically managing the resources currently assigned to a partition (virtual server), by combining the HMC DLPAR functions with partition (virtual server) properties. SDMC also allows partition (virtual server) modifications independent of the partition (virtual server) state.

Another significant enhancement is the ability of the SDMC to manage Power Systems blades. While IVM is still a choice for Power Systems blade management, clients choosing SDMC management of blades can take advantage of multiple Virtual I/O Servers, multiple shared processor pools, Active Memory Expansion, Partition (virtual server) Suspend and Resume, and many other PowerVM capabilities.

Because SDMC uses Systems Director, you can take advantage of all that Systems Director Express Edition has to offer with respect to managing Power Systems servers and blades. Highlights include:

- Operating system management: Use SDMC to discover and directly manage your operating systems.
- Event automation plans: Create automation plans that perform specific functions on a given event. For example, when the CPU utilization exceeds a given threshold, an email notification can be sent to an administrator.
- Dashboard: The SDMC provides a dashboard with health and status.
- Monitoring: Systems Director brings a richer set of monitoring functions to the management console.
- Common look and feel of the IBM Systems Director portfolio family.

The Systems Director Management Console Virtual Appliance (5765-MCV) does not include support for the customer-installed hypervisor. If you use the SDMC Virtual Appliance, you will need to purchase KVM or VMware support separately if you want to have support for your hypervisor.

- Migration, updates, and compatibility: IBM does not support upgrades or conversions from existing CR5/CR6 hardware-based Hardware Management Consoles to the Systems Director Management Console.

Differences from HMC: The following functions provided by the HMC are not provided by the SDMC: modem support and VPN call home. Also, POWER5 technology-based servers are not supported by the SDMC.

For a complete description of the possibilities offered by SDMC, see IBM Systems Director Management Console: Introduction and Overview, SG24-7860-00 available at web page:

http://www.redbooks.ibm.com/abstracts/sg247860.html?Open

2.16 Operating system support

The IBM POWER7 processor-based systems support three families of 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, and/or PowerVM Active Memory Sharing.

Note: For details about the software available on IBM POWER servers, visit the Power Systems Software™ site:

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

2.16.1 Virtual I/O Server

The minimum required level of Virtual I/O Server software depends on the server model:

Power 710 and 730	Virtual I/O Server Version 2.2, or later
Power 720 and 740	Virtual I/O Server Version 2.2, or later
Power 750	Virtual I/O Server Version 2.1.2.11 with Fix Pack 22.1 and Service Pack 1
Power 755	The Virtual I/O Server feature is not available on this model.
Power 770 and 780	Virtual I/O Server Version 2.1.2.12 with Fix Pack 22.1 and Service Pack 2
Power 795	Virtual I/O Server Version 2.2, or later

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

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

2.16.2 IBM AIX operating system

The following sections discuss the support for the various levels of AIX operating system support.

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 Web site:

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

The Fix Central Web site also provides information about ho to obtain the fixes shipping on CD-ROM.

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

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

IBM AIX Version 5.3

IBM AIX Version 5.3 is supported on all models of POWER7 processor-based servers delivered in 2010.

The minimum level of AIX Version 5.3 to support the Power 710, 720, 730, and 740 is:

- AIX 5.3 with the 5300-10 Technology Level and Service Pack 5, or later
- AIX 5.3 with the 5300-11 Technology Level and Service Pack 5, or later

- AIX 5.3 with the 5300-12 Technology Level and Service Pack 2, or later

The minimum level of AIX Version 5.3 to support the Power 750, 755, 770, and 780 is:

- AIX 5.3 with the 5300-09 Technology Level and Service Pack 7, or later
- AIX 5.3 with the 5300-10 Technology Level and Service Pack 4, or later
- AIX 5.3 with the 5300-11 Technology Level and Service Pack 2, or later

The minimum level of AIX Version 5.3 to support the Power 795 is:

- AIX 5.3 with the 5300-10 Technology Level and Service Pack 5, or later
- AIX 5.3 with the 5300-11 Technology Level and Service Pack 5, or later
- AIX 5.3 with the 5300-12 Technology Level and Service Pack 1, or later

A partition using AIX Version 5.3 will be executing in POWER6 or POWER6+ compatibility mode. This means that although the POWER7 processor has the ability to run four hardware threads per core simultaneously, using AIX 5.3 limits the number of hardware threads per core to two.

A partition with AIX 5.3 is limited to a maximum of 64 cores.

IBM AIX Version 6.1

If you install AIX 6.1 on a POWER7 processor-based server, the minimum level requirements depend on the target server model:

The minimum level of AIX Version 6.1 to support the Power 710, 720, 730, 740, and 795 is:

- AIX 6.1 with the 6100-04 Technology Level and Service Pack 7, or later
- AIX 6.1 with the 6100-05 Technology Level and Service Pack 3, or later
- AIX 6,1 with the 6100-06 Technology Level

The minimum level of AIX Version 6.1 to support the Power 750 and 755 is:

- AIX 6.1 with the 6100-02 Technology Level and Service Pack 8, or later
- AIX 6.1 with the 6100-03 Technology Level and Service Pack 5, or later
- AIX 6.1 with the 6100-04 Technology Level and Service Pack 2, or later

The minimum level of AIX Version 6.1 to support the Power 770 and 780 is:

- AIX 6.1 with the 6100-02 Technology Level and Service Pack 8, or later
- AIX 6.1 with the 6100-03 Technology Level and Service Pack 5, or later
- AIX 6.1 with the 6100-04 Technology Level and Service Pack 3, or later

A partition using AIX 6.1 with TL6 can run in POWER6, POWER6+ or POWER7 mode. It is best to run the partition in POWER7 mode to allow exploitation of new hardware capabilities such as SMT4 and Active Memory Expansion (AME).

A partition with AIX 6.1 is limited to a maximum of 64 cores.

IBM AIX Version 7.1

AIX Version 7.1 comes will full support for the Power 710, 720, 730, 740, 750, 755, 770, 780, and 795, exploiting all the hardware features from the POWER7 processor, as well as from the server architecture. A partition with AIX 7.1 can run in POWER6, POWER6+ or POWER7 mode, to enable Live Partition Mobility to different POWER6 and POWER7 systems. When

running in POWER7 mode, a partition with AIX 7.1 can scale up to 256 cores and 8 TB of RAM.

Note: Partition sizes greater than 128-cores (up to 256-cores) will require a software key to enable. Purchase will require lab services pre-analysis as a prerequisite to shipment. The software key requires feature #1256 to be installed.

2.16.3 IBM i operating system

The IBM i operating system is supported on Power 710, 720, 730, 740, 750, 770, 780, and 795 at the following minimum levels:

- IBM i Version 6.1 with i 6.1.1 machine code, or later
- IBM i Version 7.1, or later

IBM i Standard Edition, and Application Server Edition options are available the Power 740, 750, 770, 780, and 795.

- IBM i Standard Edition offers an integrated operating environment for business processing
- IBM i Application Server Edition offers IBM i without DB2 for application and infrastructure serving.

IBM i is not supported on Power 755.

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 Web site:

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

2.16.4 Linux operating system

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

The supported versions of Linux on POWER7 processor-based servers are:

- SUSE Linux Enterprise Server 10 with SP3, enabled to run in POWER6 Compatibility mode
- SUSE Linux Enterprise Server 11, supporting POWER6 or POWER7 mode
- Red Hat Enterprise Linux AP 5 Update 5 for POWER, or later

Clients wanting to configure Linux partitions in virtualized Power Systems have to be aware of these 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 may acquire Linux operating system licenses from IBM, to be included with the POWER7 processor-based servers, or from other Linux distributors.

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

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

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

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

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

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

Supported virtualization features are listed in 3.5.9, “Operating System support for PowerVM”.

2.17 Compiler technology

You can boost performance and productivity with IBM compilers on IBM Power Systems.

IBM® XL C, XL C/C++ and XL FORTRAN compilers for AIX and for Linux exploit the latest POWER7 processor architecture. Release after release, these compilers continue to help improve application performance and capability, exploiting architectural enhancements 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-to-release compiler improvements that support the POWER4 processors, through to the POWER4+, POWER5, POWER5+ and POWER6 processors, and now including the new POWER7 processors. With the support of the latest POWER7 processor chip, IBM advances a more than 20-year investment in the XL compilers for POWER series and PowerPC series architectures.

The XL C, XL C/C++ and XL FORTRAN features introduced to exploit the latest POWER7 processor include vector unit and vector scalar extension (VSX) instruction sets 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, and architecture and tuning 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. The back-end optimizer, a component common also to the IBM XL compilers, lets your applications leverage the latest industry-leading optimization technology.

PL/I for AIX supports application development on the latest POWER7 processor.

IBM Rational Development Studio for IBM i 7.1 provides programming languages for creating modern business applications. This includes the ILE RPG, ILE COBOL, C, and C++ compilers as well as the 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 new COMP-5 data type, and easier Unicode migration with relaxed USC2 rules in ILE RPG. Rational has also released a new product called Rational Open Access: RPG Edition. This opens up the ILE RPG file I/O processing, enabling partners, tool providers, and users to write custom I/O handlers that can access other devices such as 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.

IBM Rational Power Appliance is a preconfigured application development environment solution for Power Systems, which 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 Rational Team Concert for Power Systems Software, 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.

2.18 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 exploits 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.

2.18.1 IBM EnergyScale technology

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

On 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 to interact with the IBM Systems Director Active Energy Manager (an IBM Systems Directors plug-in) for energy management and control.

IBM EnergyScale makes use of power and thermal information collected from the system in order to implement policies that can lead to better performance, or better energy utilization. IBM EnergyScale features include:

Power trending

EnergyScale provides continuous collection of real-time server energy consumption. This 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 datacenter 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.

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 up to 30% down from nominal frequency (the actual value depends on the server type and configuration). Power Saver Mode is not supported during boot or re-boot although it is a persistent condition that will be 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 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 will lower the frequency and voltage to Power Energy Saver Mode values. When fully utilized, the maximum frequency will vary, 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 95% of nominal values. If performance is favored over energy consumption, the maximum frequency can be increased to up to 109% of nominal frequency for extra performance.

Dynamic Power Saver Mode is mutually exclusive with Power Saver mode. Only one of these modes can be enabled at a given time.

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 actually 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 frees up extra power 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 datacenter. 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; one is Power Capping as described before, and the other is 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

The IBM 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 is very small, typically not generating any impact on applications running. Because of that, the POWER Hypervisor can use the 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 inside 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 from 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 called Sleep. Before a core and its associated L2 and L3 caches enter Sleep mode, caches are flushed and transition lookaside buffers (TLB) are invalidated, and hardware clock is turned off in the core and in the caches. Voltage is reduced in order to minimize leakage current. Processor cores inactive in the system (such as CoD processor cores) are kept in Sleep mode. Sleep mode saves about 35% power consumption in the processor core and associated L2 and L3 caches.

Fan control and altitude input

System firmware will dynamically adjust 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.

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 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 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 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. It makes sense to do this when there will be long periods of low utilization, such as weekends. AEM provides information, such as the power that will be saved and the time it will take 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 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 newer, and POWER7 systems with EM730 firmware or newer, is the capability to set a power savings mode for partitions or the system

processor pool. As ith 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 will have a common power savings setting, which is that of the system processor pool. This 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, or favor partition power savings. The user must configure this setting 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 processor cores that belong to the partition. Based on this utilization and performance data, the firmware will dynamically adjust the processor frequency and voltage, reacting within milliseconds to adjust workload performance and also deliver power savings when the partition is under-utilized.

In addition to the two Dynamic Power Saver options, the customer can select to have no power savings on a given partition. This option will leave the processor cores assigned to the partition running at their nominal frequencies and voltages.

A new power savings mode, called “Inherit Host Setting”, is available and is only applicable to partitions. When configured to use this setting, a partition will adopt the power savings mode of its hosting server. By default, all partitions with dedicated processing units, and the system processor pool, are set to Inherit Host Setting.

On POWER7 processor-based systems, several EnergyScales 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 2-20 provides a list of all features actually supported, underlining all cases where AEM is not required. The table also details the features that can be activated by traditional user interfaces (ASMI, HMC).

Table 2-20 AEM support

Feature	Active Energy Manager (AEM) required	ASMI	HMC
Power Trending	Y	N	N
Thermal Reporting	Y	N	N
Static Power Saver	N	Y	Y
Dynamic Power Saver	Y	N	N
Power Capping	Y	N	N
Energy-optimized Fans	N	-	-
Processor Core Nap	N	-	-
Processor Core Sleep	N	-	-
Processor Folding	N	-	-
EnergyScale for I/O	N	-	-
Server Power Down	Y	-	-
Partition Power Management	Y	-	-

The Power 710 and Power 730 implement all the Energy Scale capabilities listed in 2.18.1, “IBM EnergyScale technology” on page 70.

2.18.2 Thermal power management device card

The thermal power management device (TPMD) card is a separate microcontroller installed on some POWER6 processor-based systems, and in all 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; and 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.

For information about IBM EnergyScale technology, go to:

http://www.ibm.com/common/ssi/cgi-bin/ssialias?infotype=SA&subtype=WH&appname=STGE_PO_PO_USEN&htmlfid=POW03039USEN&attachment=POW03039USEN.PDF

3

Virtualization

As you look for ways to maximize the return on your IT infrastructure investments, consolidating workloads becomes an attractive proposition.

IBM Power Systems combined with PowerVM technology are designed to help you consolidate and simplify your IT environment, with the following key capabilities:

- 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, thereby enabling you to make business-driven policies to deliver resources based on time, cost, and service-level requirements.

This chapter discusses the virtualization technologies and features on IBM Power Systems:

- POWER Hypervisor
- POWER Modes
- Partitioning
- Active Memory Expansion
- PowerVM
- System Planning Tool

3.1 POWER Version 2.2 enhancements

The latest available PowerVM Version 2.2 contains the following enhancements:

- Support for up to 80 LPARs on Power 710 and 720
- Support for up to 160 LPARs on Power 730, 740,
- Support for up to 320 LPARs on Power 750,
- Support for up to 640 LPARs on Power 770 and 780
- Support for up to 1024 LPARs on Power 795
- PowerVM support for sub-chip per-core licensing on Power 710, 720, 730, and 740
- Role Based Access Control (RBAC):

  RBAC brings an added level of security and flexibility in the administration of VIOS. With RBAC, you can create a set of authorizations for the user management commands. You can assign these authorizations to a role UserManagement and this role can be given to any other user. So a normal user with the role UserManagement can manage the users on the system but will not have any further access.

  With RBAC, the Virtual I/O Server has the capability of split management functions that presently can be done only by the *padmin* user, providing better security by giving only the necessary access to users, and easy management and auditing of system functions.
- Support for Concurrent Add of VLANs
- Support for USB tape:

  The Virtual I/O Server now supports a USB DAT-320 Tape Drive and its use as a virtual tape device for VIOS clients.
- Support for USB Blu-ray:

  The Virtual I/O Server now supports USB Blu-ray optical devices. AIX does not support mapping these as virtual optical devices to clients. However, you can import the disk into the virtual optical media library and map the created file to the client as a virtual DVD drive.

The IBM PowerVM Workload Partitions Manager™ for AIX, Version 2.2 has the following enhancements:

- When used with AIX 6.1 Technology Level 6, the following support applies:
  - Support for exporting VIOS SCSI disk into a WPAR. Compatibility analysis and mobility of WPARs with VIOS SCSI disk. In addition to Fibre Channel devices, now VIOS SCSI disks can be exported into a workload partition (WPAR).
  - WPAR Manager Command-Line Interface (CLI). The WPAR Manager CLI allows federated management of WPARs across multiple systems by command line.
  - Support for workload partition definitions. The WPAR definitions can be preserved after WPARs are deleted. These definitions can be deployed at a later time to any WPAR-capable system.
- In addition to the features supported on AIX 6.1 Technology Level 6, the following support applies to AIX 7.1:
  - Support for AIX 5.2 Workload Partitions for AIX 7.1. Life cycle management and mobility enablement for AIX 5.2 Technology Level 10 SP8 Version WPARs.
  - Support for trusted kernel extension loading and configuration from WPARs. Enables exporting a list of kernel extensions that can then be loaded inside a WPAR, yet maintaining isolation.

3.2 POWER Hypervisor

Combined with features designed into the POWER7 processors, the POWER Hypervisor delivers functions that enable other system technologies, including logical partitioning technology, virtualized processors, IEEE VLAN compatible virtual switch, virtual SCSI adapters, 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 87)
- Saves and restores all processor state information during a logical processor context switch
- Controls hardware I/O interrupt management facilities for logical partitions
- Provides virtual LAN channels between logical partitions that help to reduce the need for physical Ethernet adapters for inter-partition communication
- Monitors the Service Processor and will perform a reset/reload if it detects the loss of the Service Processor, notifying the operating system if the problem is not corrected

The POWER Hypervisor is always active, regardless of the system configuration and also when not connected to the HMC. It requires memory to support the resource assignment to the logical partitions on the server. The amount of memory required by the POWER Hypervisor firmware varies according to various factors. Here we list factors influencing the POWER Hypervisor memory requirements:

- 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 to create a partition is the size of the system's Logical Memory Block (LMB). The default LMB size varies according to the amount of memory configured in the CEC as shown in Table 3-1.

Table 3-1 Configured CEC memory-to-default Logical Memory Block size

Configurable CEC memory	Default Logical Memory Block
Greater than 8 GB up to 16 GB	64 MB
Greater than 16 GB up to 32 GB	128 MB
Greater than 32 GB	256 MB

But in most cases, the actual minimum requirements and preferable capacities of the supported operating systems are above 256 MB. Physical memory is assigned to partitions in increments of Logical Memory Block.

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

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

3.2.1 Virtual SCSI

The POWER Hypervisor provides a virtual SCSI mechanism for virtualization of storage devices. The storage virtualization is accomplished using two paired adapters: a virtual SCSI server adapter and 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 3.5.4, "Virtual I/O Server" on page 93. The Virtual I/O Server software is available with the optional PowerVM Edition features.

3.2.2 Virtual Ethernet

The POWER Hypervisor provides a virtual Ethernet switch function that allows partitions on the same server to use a fast and secure communication without any need for physical interconnection. The virtual Ethernet allows a transmission speed in the range of 1 to 3 Gbps. depending on the MTU[1] size and CPU entitlement. Virtual Ethernet support starts with AIX Version 5.3, or appropriate level of Linux supporting Virtual Ethernet devices (see 3.5.9, "Operating System support for PowerVM" on page 100). The virtual Ethernet is part of the base system configuration.

Virtual Ethernet has the following major features:

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

Any virtual Ethernet can also have connectivity outside of the server if a layer-2 bridge to a physical Ethernet adapter is set in one Virtual I/O server partition (see 3.5.4, "Virtual I/O Server" on page 93 for more details about shared Ethernet). This is also known as a Shared Ethernet Adapter.

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

[1] Maximum transmission unit

3.2.3 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 3.5.8, “NPIV” on page 100.

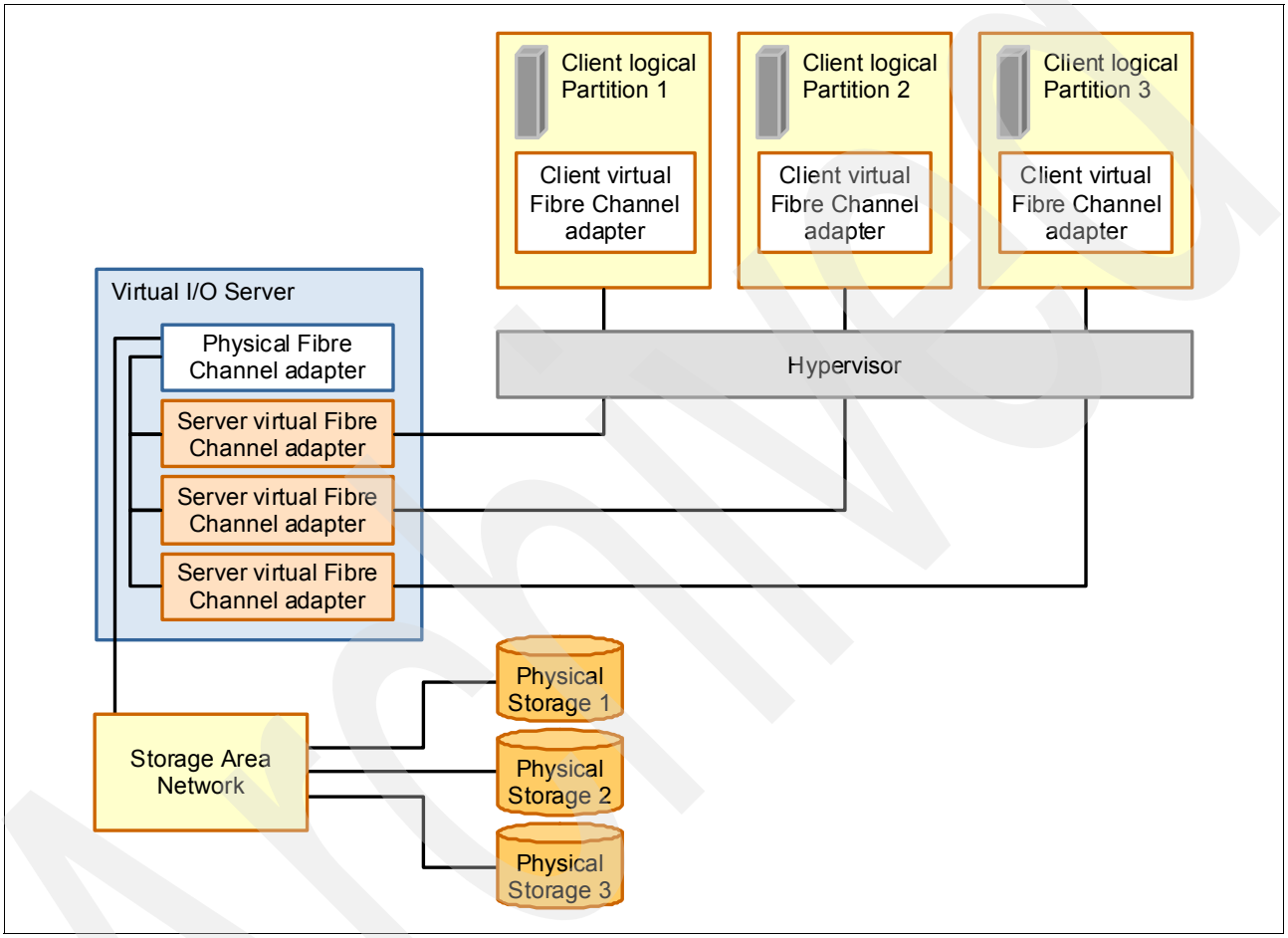


Figure 3-1 Connectivity between virtual Fibre Channels adapters and external SAN devices

3.2.4 Virtual (TTY) console

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

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

3.3 POWER processor modes

Although not a virtualization feature, strictly speaking, POWER modes are described here because they have an impact on certain virtualization features.

On any Power Systems server, partitions can be configured to run in various modes:

- POWER6 compatibility mode:

  This execution mode is compatible with Version 2.05 of the Power Instruction Set Architecture (ISA).

  http://www.power.org/resources/reading/PowerISA_V2.05.pdf

- POWER6+ compatibility mode:

  This mode is similar to the POWER6 compatibility mode, with 8 additional Storage Protection Keys.

- POWER7 mode:

  This is the native mode for POWER7 processors, implementing the v2.06 of the Power Instruction Set Architecture.

  http://www.power.org/resources/downloads/PowerISA_V2.06_PUBLIC.pdf

The selection of the mode is made on a per partition basis, from the HMC, by editing the partition profile as presented in Figure 3-2.

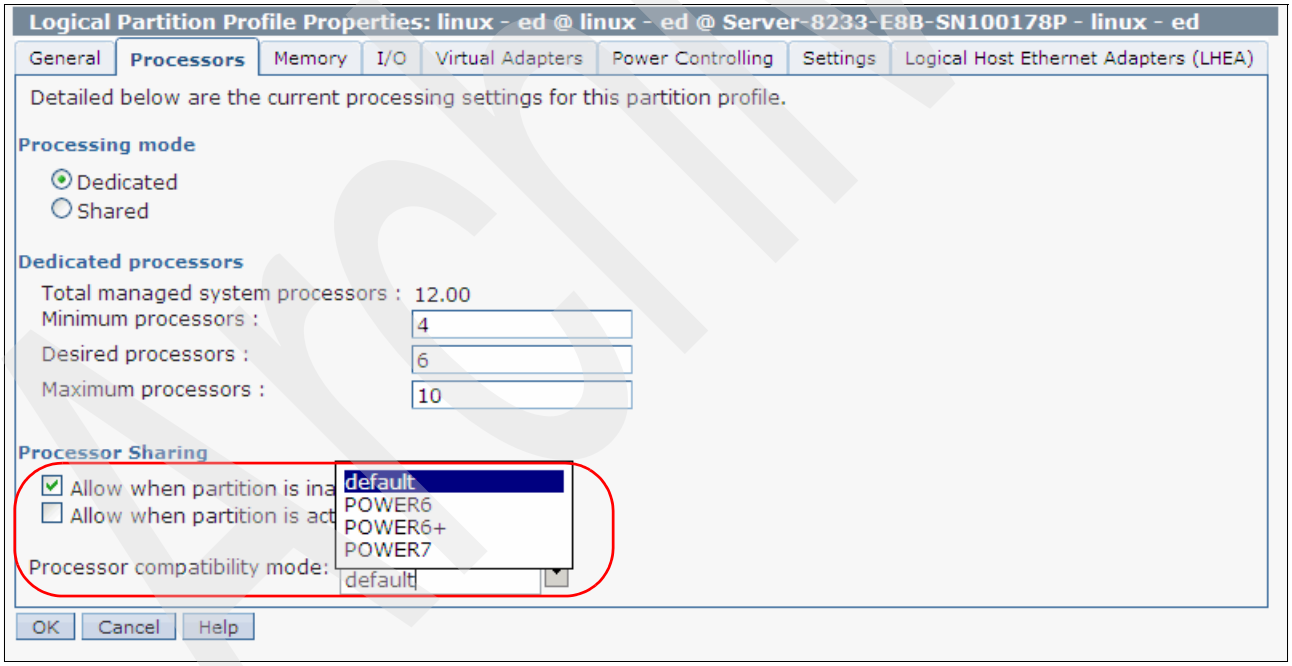


Figure 3-2 Configuring partition profile compatibility mode from the HMC

Table 3-2 lists the differences between these modes.

Table 3-2 Differences between POWER6 and POWER7 mode

POWER6 mode (and POWER6+)	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	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/128-thread Scaling	► 32-core / 128-thread Scaling ► 64-core / 256-thread Scaling ► 256-core / 1024-thread Scaling	Performance and Scalability for Large Scale-Up Single System Image Workloads (for example, OLTP, ERP scale-up, WPAR consolidation)
EnergyScale CPU Idle	EnergyScale CPU Idle and Folding with NAP and SLEEP	Improved Energy Efficiency

3.4 Active Memory Expansion

Active Memory Expansion is an innovative POWER7 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 100%. This can allow a partition to do 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 Enablement is an optional hardware feature of POWER7 processor-based servers, which must be specified when creating the configuration in the e-config tool. Active Memory Expansion is enabled by adding the hardware feature #4795.

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

Active Memory Expansion uses CPU resource of a partition to compress/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/decompression. 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.

Clients have a great deal of 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 desired in each partition to help control the amount of CPU used by the Active Memory Expansion function. An IPL is required for the specific partition that is turning memory expansion on or off. When turned on, there are monitoring capabilities in standard AIX performance tools such as **lparstat**, **vmstat**, **topas**, and **svmon**.

Figure 3-3 represents the percentage of CPU used to compress memory for two partitions with various profiles. The green curve corresponds to a partition that has spare processing power capacity, while the blue curve corresponds to a partition constrained in processing power.

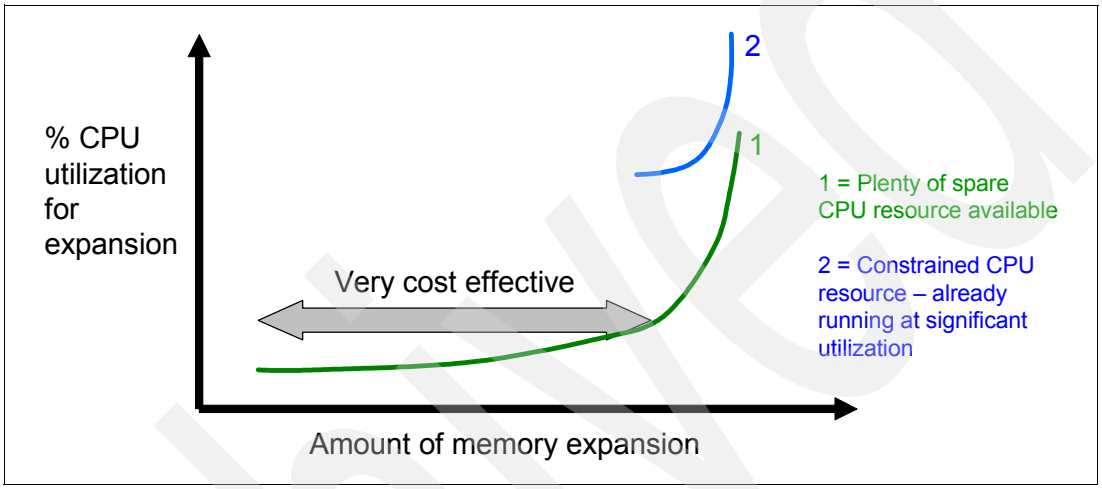


Figure 3-3 CPU usage versus memory expansion effectiveness

Both cases show that there is a knee-of-curve relationship for 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 memory contents are. This demonstrates the need for a case by case study of whether memory expansion can provide a positive return on investment.

To help you perform this study, a planning tool is included with AIX 6.1 Technology Level 4 allowing you to sample actual workloads and estimate both 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 returned by this planning tool. The tool outputs various real memory and CPU resource combinations to achieve the desired effective memory and indicates one particular combination. In this example, the tool proposes to allocate 58% of a processor core, to benefit from 45% extra memory capacity.

```
Active Memory Expansion Modeled Statistics:
-----------------------
Modeled Expanded Memory Size :   8.00 GB

Expansion      True Memory       Modeled Memory        CPU Usage
 Factor        Modeled Size      Gain                  Estimate
---------    --------------    -----------------    -----------
  1.21          6.75 GB          1.25 GB [ 19%]        0.00
  1.31          6.25 GB          1.75 GB [ 28%]        0.20
  1.41          5.75 GB          2.25 GB [ 39%]        0.35
  1.51          5.50 GB          2.50 GB [ 45%]        0.58
  1.61          5.00 GB          3.00 GB [ 60%]        1.46

Active Memory Expansion Recommendation:
---------------------
The recommended AME configuration for this workload is to configure
the LPAR with a memory size of 5.50 GB and to configure a memory
expansion factor of 1.51.  This will result in a memory expansion of
45% from the LPAR's current memory size.  With this configuration,
the estimated CPU usage due to Active Memory Expansion is
approximately 0.58 physical processors, and the estimated overall
peak CPU resource required for the LPAR is 3.72 physical processors.
```

Figure 3-4 Output form Active Memory Expansion planning tool

When you have selected the value of the memory expansion factor you want to achieve, you can use this value to configure the partition from the HMC, as shown in Figure 3-5.

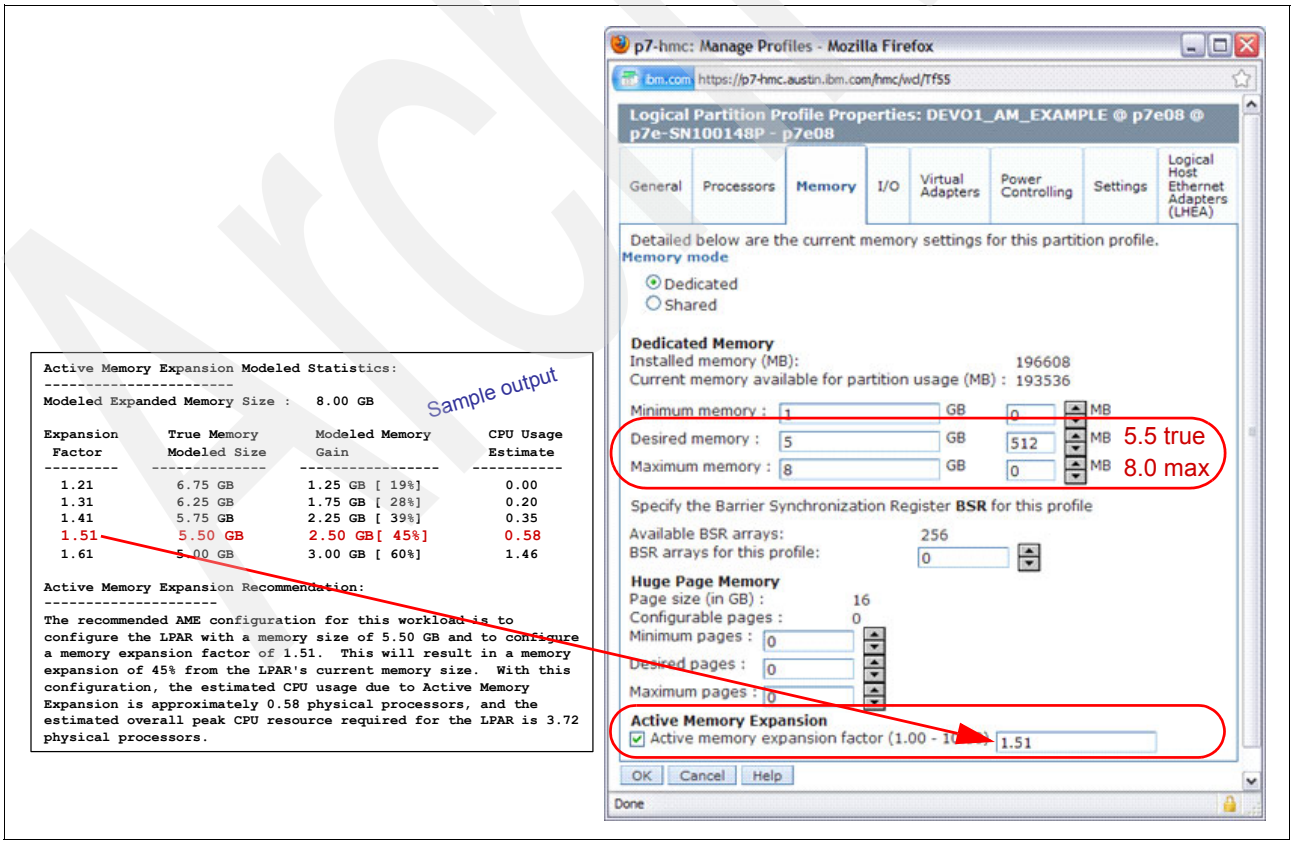


Figure 3-5 Using the planning tool results to configure the partition

On the HMC menu describing the partition, select the Active Memory Expansion check box and enter true and max memory, as well as the memory expansion factor. To turn off expansion, you just have to unselect the check box. In both cases, a reboot of the partition is needed 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 using the 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 an MES order. A software key is provided when the enablement feature is ordered that is applied to the server. A reboot is not required to enable the physical server. The key is specific to an individual server and is permanent. It cannot be moved to another server. This feature is ordered per server, independently of the number of partitions using memory expansion.

It is possible to view if the Active Memory Expansion feature has been activated on a server from the HMC, as presented in Figure 3-6.

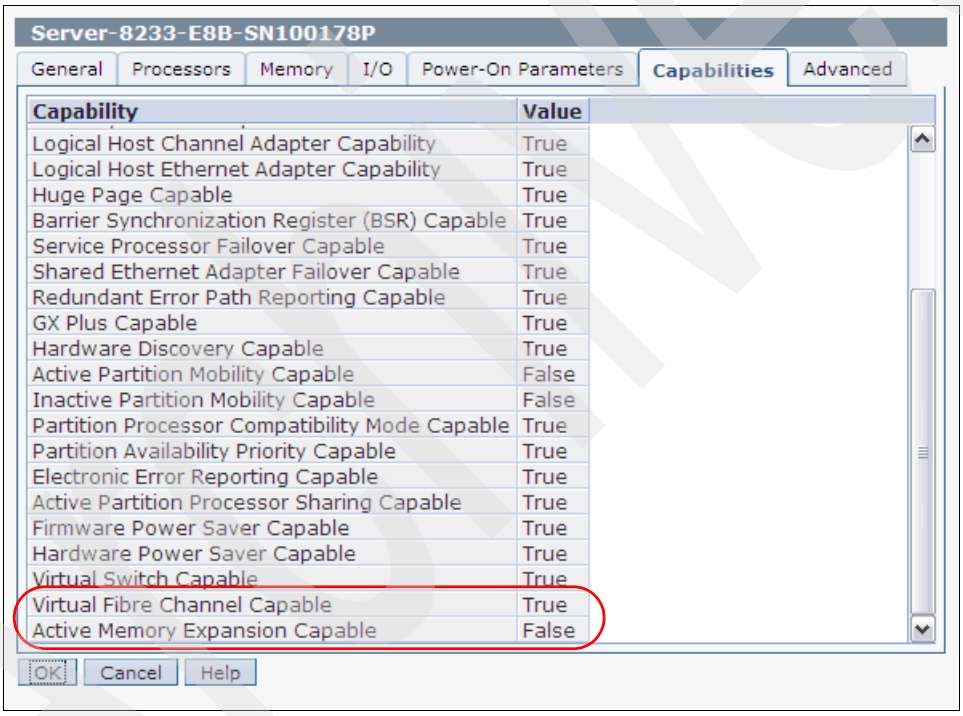


Figure 3-6 Server capabilities listed from the HMC

Note: In order to move an LPAR using Active Memory Expansion through Live Partition Mobility to another system, the target system must support AME. That means the target system must have AME activated through the software key. If the target system does not have AME activated, the mobility operation will fail during the pre-mobility check phase, and an appropriate error message will be displayed to the user.

For detailed information regarding Active Memory Expansion, you can download the document *Active Memory Expansion: Overview and Usage Guide* from

http://www.ibm.com/common/ssi/fcgi-bin/ssialias?infotype=SA&subtype=WH&appname=STGE_PO_PO_USEN&htmlfid=POW03037USEN&attachment=POW03037USEN.PDF

3.5 PowerVM

PowerVM is the industry-leading virtualization solution for AIX, IBM i, and Linux environments on IBM POWER processor-based systems. PowerVM offers a secure virtualization environment, built on the advanced RAS features and leadership performance of the Power Systems platform. PowerVM features leading technologies such as Logical Partitioning, IBM Micro-Partitioning™, Power Hypervisor, Virtual I/O Server, PowerVM Lx86, PowerVM Live Partition Mobility, and PowerVM Active Memory Sharing. As with the previous Advanced Power Virtualization solution, PowerVM is a combination of hardware enablement and value-added software. The licensed features of each of the three editions of PowerVM are discussed in 3.5.1, “PowerVM editions” on page 85.

3.5.1 PowerVM editions

This section provides information about the virtualization capabilities of PowerVM. There are three versions of PowerVM, which are suited for various purposes:

- PowerVM Express Edition:

  PowerVM Express Edition is designed for customers looking for an introduction to more advanced virtualization features at a highly affordable price.

- PowerVM Standard Edition:

  PowerVM Standard Edition provides the most complete virtualization functionality for AIX, IBM i, and Linux operating systems in the industry. PowerVM Standard Edition is supported on all POWER processor-based servers and includes features designed to allow businesses to increase system utilization.

- PowerVM Enterprise Edition:

  PowerVM Enterprise Edition includes all the features of PowerVM Standard Edition plus two new industry-leading capabilities called Active Memory Sharing and Live Partition Mobility.

Table 3-3 lists the feature codes of PowerVM editions that are available on the Power 710 and Power 730 servers.

Table 3-3 Availability of PowerVM editions on the Power 710 and Power 730

PowerVM Editions	Express	Standard	Enterprise
IBM Power 710	#5225	#5227	#5228
IBM Power 730	#5225	#5227	#5228

It is possible to upgrade from the Express Edition to the Standard or Enterprise Edition, and from Standard to Enterprise Editions. Table 3-4 outlines the functional elements of the three PowerVM editions.

Table 3-4 PowerVM capabilities

PowerVM Editions	Express	Standard	Enterprise
Micro-partitions	Yes	Yes	Yes
Maximum LPARs	1 or 2 per server	10/core	10/core
Management	IVM	IVM or HMC	IVM or HMC
Virtual IO Server	Yes	Yes	Yes

PowerVM Editions	Express	Standard	Enterprise
NPIV	Yes	Yes	Yes
Multiple Shared Processor Pool	No	Yes	Yes
Live Partition Mobility	No	No	Yes
Active Memory Sharing	No	No	Yes

3.5.2 Logical partitions

Logical partitions (LPARs) and virtualization increase utilization of system resources and add a new level of configuration possibilities. This section provides details and configuration specifications for these topics.

Dynamic logical partitioning

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

Later, dynamic 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. IBM AIX Version 5.2, with all the necessary enhancements to enable dynamic LPARs, was introduced in 2002. The ability to reconfigure dynamic LPARs encourages system administrators to dynamically redefine all available system resources to reach the optimum capacity for each defined dynamic LPAR.

Micro-Partitioning

Micro-Partitioning technology allows you to allocate fractions of processors to a logical partition. This technology was introduced with POWER5 processor-based systems. A logical partition using fractions of processors is also known as a Shared Processor Partition or Micro-Partition. Micro-Partitions run over a set of processors called Shared Processor Pool.

Virtual processors are used to let the operating system manage the fractions of processing power assigned to the logical partition. From an operating system perspective, a virtual processor cannot be distinguished from a physical processor, unless the operating system has been enhanced to be made aware of the difference. Physical processors are abstracted into virtual processors that are available to partitions. The meaning of the term *physical processor* in this section is a *processor core*. For example, in a 2-core server there are two physical processors.

When defining a shared processor partition, various options have to 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. Pick one from the list with the names of each configured shared processor pool. This list also displays the pool ID of each configured shared processor pool in parentheses. If the name of the desired shared processor pool is not available here, you must first configure the desired shared processor pool using the Shared Processor Pool Management window. Shared processor partitions use the default shared processor pool called DefaultPool by default.

- Whether the partition will be able or not to access extra processing power to “fill up” its virtual processors above its capacity entitlement. You select either to cap or uncap your partition. If there is spare processing power available in the shared processor pool, or if 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’s processing power based on minimum, desired, and maximum values, processing mode, and also based on other active partitions’ requirements. The actual entitlement is never smaller than the processing unit’s desired value but can exceed that value in the case of an uncapped partition and up to the number of virtual processors allocated.

A partition can be defined with a processor capacity as small as 0.10 processing units. This represents 1/10th of a physical processor. Each physical processor can be shared by up to ten shared processor partitions, and the partition’s entitlement can be incremented fractionally by as little as 1/100th of the processor. The shared processor partitions are dispatched and time-sliced on the physical processors under control of the POWER Hypervisor. The shared processor partitions are created and managed by the HMC or Integrated Virtualization Management.

The IBM Power 710 system can be configured with up to eight cores, and the IBM Power 730 up to 16 cores. These systems can support LPARs as follows:

- Up to 80 LPARs on a Power 710
- Up to 160 LPARs on a Power 730

Important: The maximums stated are supported by the hardware, but the practical limits depend on the application workload demands.

Here we provide additional information about virtual processors:

- A virtual processor can be either running (dispatched) on a physical processor or standby waiting for a physical processor to became available.
- Virtual processors do not introduce any additional abstraction level; they really are only a dispatch entity. When running on a physical processor, virtual processors run at the same speed as the physical processor.
- Each partition’s profile defines CPU entitlement that determines how much processing power any given partition must 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 processor units from a shared processor pool. This setting will define the processing mode of the logical partition. Figure 3-7 shows a diagram of the concepts discussed in this section.

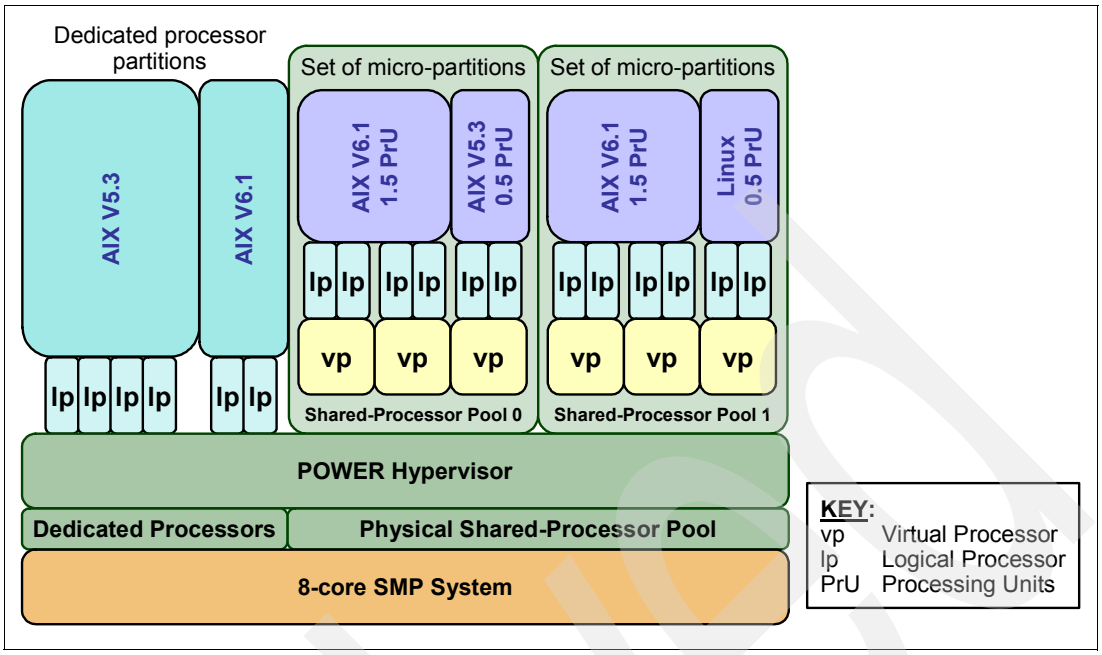


Figure 3-7 Logical partitioning concepts

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, we use 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). If simultaneous multithreading is off, then 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 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 assigned to the partition). From the POWER Hypervisor's point of view, 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 must receive total CPU time equal to their processing units 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 on whether the simultaneous multithreading is turned on or off.

3.5.3 Multiple Shared-Processor Pools

Multiple Shared-Processor Pools (MSPPs) is a capability supported on POWER7 and POWER6 processor-based servers. This capability allows a system administrator to create a set of micro-partitions with the purpose of controlling the processor capacity that can be consumed from the physical shared-processor pool.

To implement MSPPs, there is a set of underlying techniques and technologies. An overview of the architecture of Multiple Shared-Processor Pools can be seen in Figure 3-8.

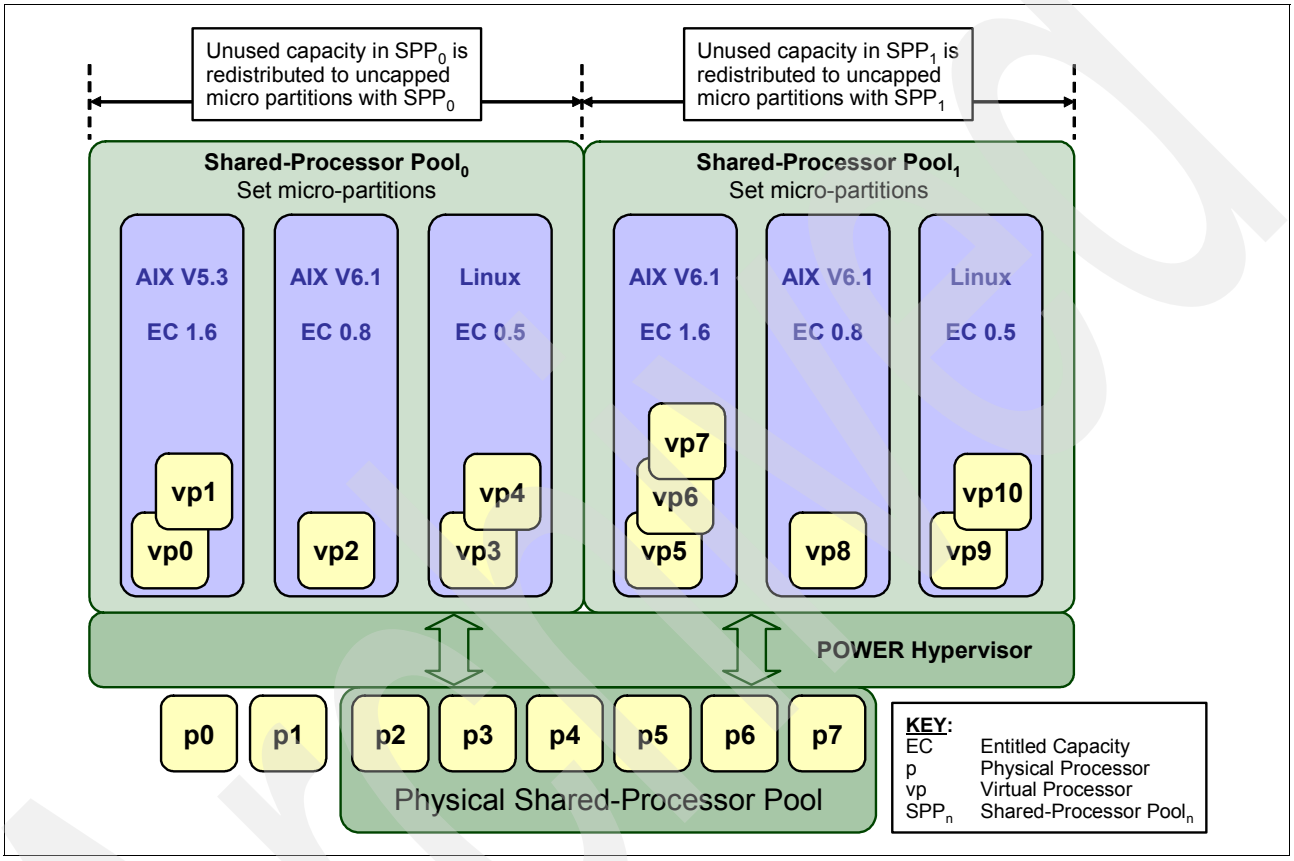


Figure 3-8 Overview of the architecture of Multiple Shared-Processor Pools

Micro-partitions 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 micro-partitions 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 micro-partition within a Shared-Processor Pool is guaranteed its processor entitlement plus any capacity that it can be allocated from the Reserved Pool Capacity if the micro-partition is uncapped.

If certain micro-partitions in a Shared-Processor Pool do not use their capacity entitlement, the unused capacity is ceded and other uncapped micro-partitions 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 micro-partitions within that Shared-Processor Pool.

All Power Systems servers that support the Multiple Shared-Processor Pools capability will 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 3-5 Attribute values for the default Shared-Processor Pool (SPP_0)

SPP_0 attribute	**Value**
Shared-Processor Pool ID	0
Maximum Pool Capacity	The 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 micro-partitions 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 micro-partitions (either transferred from SPP_0 or newly created).

Levels of processor capacity resolution

There are two levels of processor capacity resolution implemented by the POWER Hypervisor and Multiple Shared-Processor Pools:

$Level_0$ The first level, $Level_0$, 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 micro-partition within the same Shared-Processor Pool.

$Level_1$ 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 micro-partitions regardless of the Multiple Shared-Processor Pools structure. This is the second level of processor capacity resolution.

You can see the two levels of unused capacity redistribution implemented by the POWER Hypervisor in Figure 3-9.

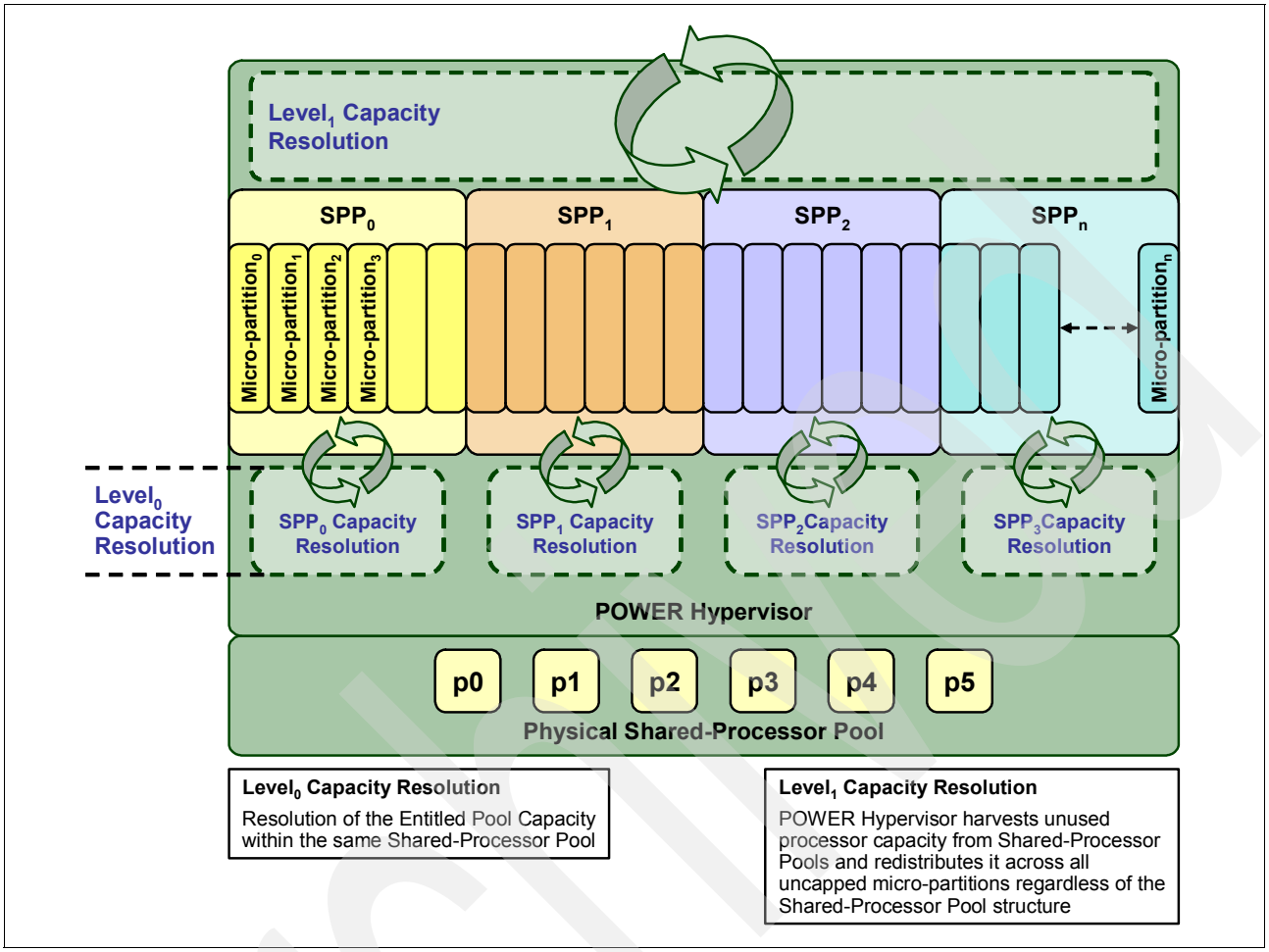


Figure 3-9 The two levels of unused capacity redistribution

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 is where unused processor capacity within a Shared-Processor Pool is harvested and then redistributed to uncapped micro-partitions 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 a few of the uncapped micro-partitions within it require additional processor capacity (above the Entitled Pool Capacity) then the POWER Hypervisor redistributes part of the extra capacity to the uncapped micro-partitions. This level of processor capacity management is sometimes referred to as Level$_1$ capacity resolution.

To redistribute unused processor capacity to uncapped micro-partitions in Multiple Shared-Processor Pools above the Entitled Pool Capacity, the POWER Hypervisor uses a higher level of redistribution, Level$_1$.

Important: For $Level_1$ capacity resolution, when allocating additional processor capacity in excess of the Entitled Pool Capacity of the Shared-Processor Pool, the POWER Hypervisor takes into account the uncapped weights of *all micro-partitions in the system, regardless of the Multiple Shared-Processor Pool structure*.

Where there is unused processor capacity in underutilized Shared-Processor Pools, the micro-partitions within the Shared-Processor Pools cede the capacity to the POWER Hypervisor.

In busy Shared-Processor Pools where the micro-partitions have used all of the Entitled Pool Capacity, the POWER Hypervisor will allocate additional cycles to micro-partitions under the following circumstances:

- The Maximum Pool Capacity of the Shared-Processor Pool hosting the micro-partition has not been met.
- (and) The micro-partition is uncapped.
- (and) The micro-partition has enough virtual-processors to take advantage of the additional capacity.

Under such circumstances, the POWER Hypervisor allocates additional processor capacity to micro-partitions on the basis of their uncapped weights independent of the Shared-Processor Pool hosting the micro-partitions. This 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 micro-partitions in the system into account, regardless of the Multiple Shared-Processor Pools 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 HMC using either the graphical or CLI interface.

Dynamic adjustment of Reserve 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 HMC using either the graphical or CLI interface.

Dynamic movement between Shared-Processor Pools

A micro-partition can be moved dynamically from one Shared-Processor Pool to another using the HMC by either the graphical or CLI interface. Because the Entitled Pool Capacity is partly made up of the sum of the entitled capacities of the micro-partitions, removing a micro-partition from a Shared-Processor Pool will reduce the Entitled Pool Capacity for that Shared-Processor Pool. Similarly, the Entitled Pool Capacity of the Shared-Processor Pool that the micro-partition joins will increase.

Deleting a Shared-Processor Pool

Shared-Processor Pools cannot be deleted from the system, but they can be deactivated by setting the Maximum Pool Capacity and the Reserved Pool Capacity to zero. By doing so, the Shared-Processor Pool will still exist but will not be active. You can use the HMC interface to deactivate a Shared-Processor Pool. Be aware that a Shared-Processor Pool cannot be deactivated unless all micro-partitions hosted by the Shared-Processor Pool have been removed.

Live Partition Mobility and Multiple Shared-Processor Pools

A micro-partition can leave a Shared-Processor Pool due to PowerVM Live Partition Mobility. Similarly, a micro-partition 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 micro-partition.

Because various simultaneous micro-partition migrations are supported by PowerVM Live Partition Mobility, it is conceivable to migrate the entire Shared-Processor Pool from one server to another.

3.5.4 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.

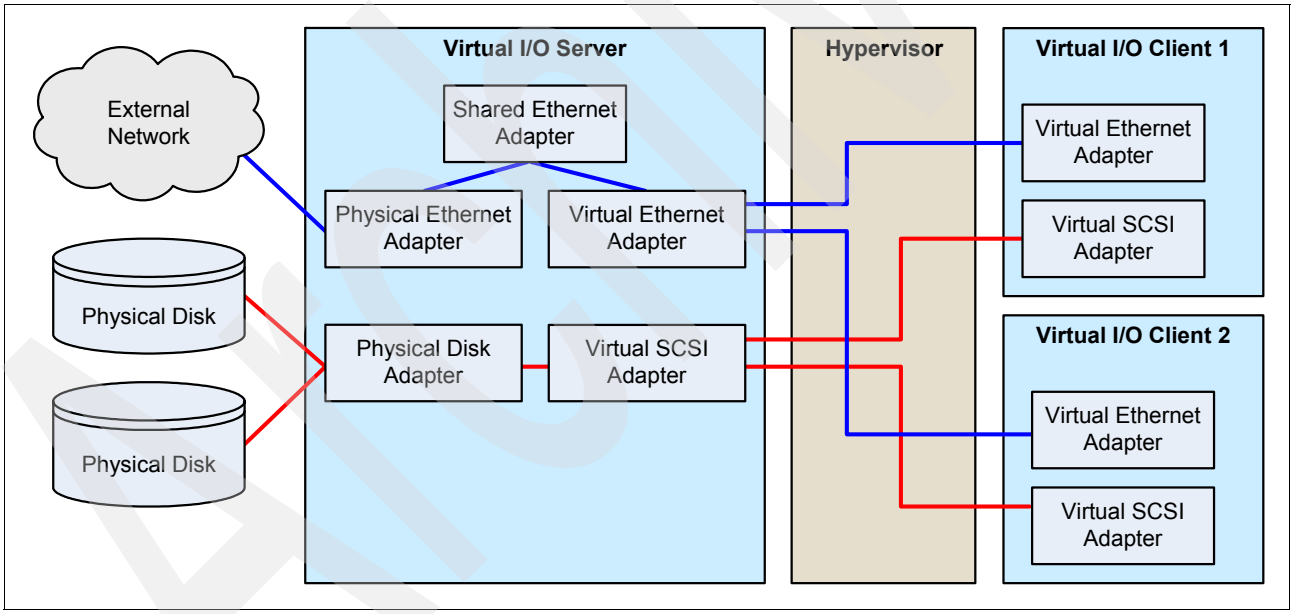


Figure 3-10 Architectural view of the Virtual I/O Server

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 that order any PowerVM edition. This dedicated software is only for the Virtual I/O Server (and IVM in case it is used) and is only supported in special Virtual I/O Server partitions. Three major virtual devices are supported by the Virtual I/O Server: a Shared Ethernet Adapter, Virtual SCSI, and Virtual Fibre Channel adapter. The Virtual Fibre Channel adapter is used with the NPIV feature, described in 3.5.8, “NPIV” on page 100.

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 various client partitions to share one physical adapter. Using an SEA, you can connect internal and external VLANs using a physical adapter. The Shared Ethernet Adapter service can only be hosted in the Virtual I/O Server, not in a general purpose AIX, IBM i, or Linux partition, and acts as a layer-2 network bridge to securely transport network traffic between virtual Ethernet networks (internal) and one or more (EtherChannel) physical network adapters (external). These virtual Ethernet network adapters are defined by the POWER Hypervisor on the Virtual I/O Server

Tip: A Linux partition can provide bridging function as well, by using the `brctl` command.

Figure 3-11 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.

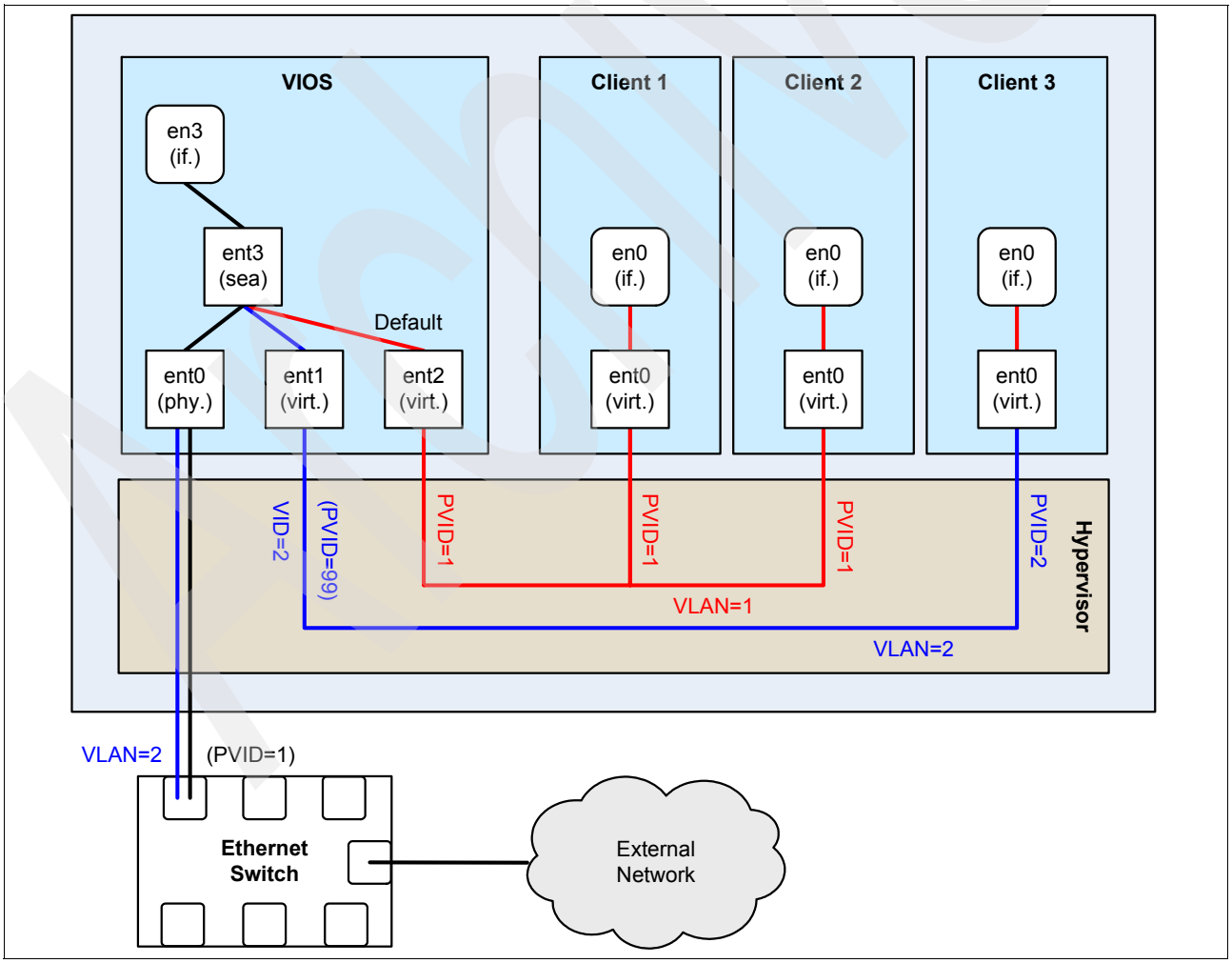


Figure 3-11 Architectural view of a Shared Ethernet Adapter

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, it is possible for a single physical Ethernet to be shared between 320 internal VLANs. The number of shared Ethernet adapters that can be set up in a Virtual I/O server partition is limited only by the resource availability, because there are no configuration limits.

Unicast, broadcast, and multicast is supported, so protocols that rely on broadcast or multicast, such as Address Resolution Protocol (ARP), Dynamic Host Configuration Protocol (DHCP), Boot Protocol (BOOTP), and Neighbor Discovery Protocol (NDP) can work across an SEA.

Note: A Shared Ethernet Adapter does not need to have an IP address configured to be able to perform the Ethernet bridging functionality. It is very convenient to configure IP on the Virtual I/O Server, 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. These operations 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, thus leaving the SEA without the IP address, and allowing for maintenance on the SEA without losing IP connectivity in case SEA failover is configured.

Virtual SCSI

Virtual SCSI is used to refer to a virtualized implementation of the SCSI protocol. Virtual SCSI is based on a client/server relationship. The Virtual I/O Server logical partition owns the physical resources and acts as server or, in SCSI terms, 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 (a virtual SCSI server adapter and a virtual SCSI client adapter) are configured using an HMC or through the Integrated Virtualization Manager. The virtual SCSI server (target) adapter is responsible for executing any SCSI commands it receives. It is owned by the Virtual I/O Server partition. The virtual SCSI client adapter allows a client partition to access physical SCSI and SAN attached devices and LUNs that are assigned to the client partition. The provisioning of virtual disk resources is provided by the Virtual I/O Server.

Physical disks presented to the Virtual/O Server can be exported and assigned to a client partition in a number of various ways:

- The entire disk is presented to the client partition.
- The disk is divided into various logical volumes, which can be presented to a single client or multiple clients.
- As of Virtual I/O Server 1.5 or later, files can be created on these disks and file backed storage devices can be created.

The logical volumes or files can be assigned to various partitions. Therefore, virtual SCSI enables sharing of adapters as well as 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 of the two client partitions is assigned one logical volume, which is then accessed through a virtual I/O adapter (VSCSI Client Adapter). Inside the partition, the disk is seen as a normal hdisk.

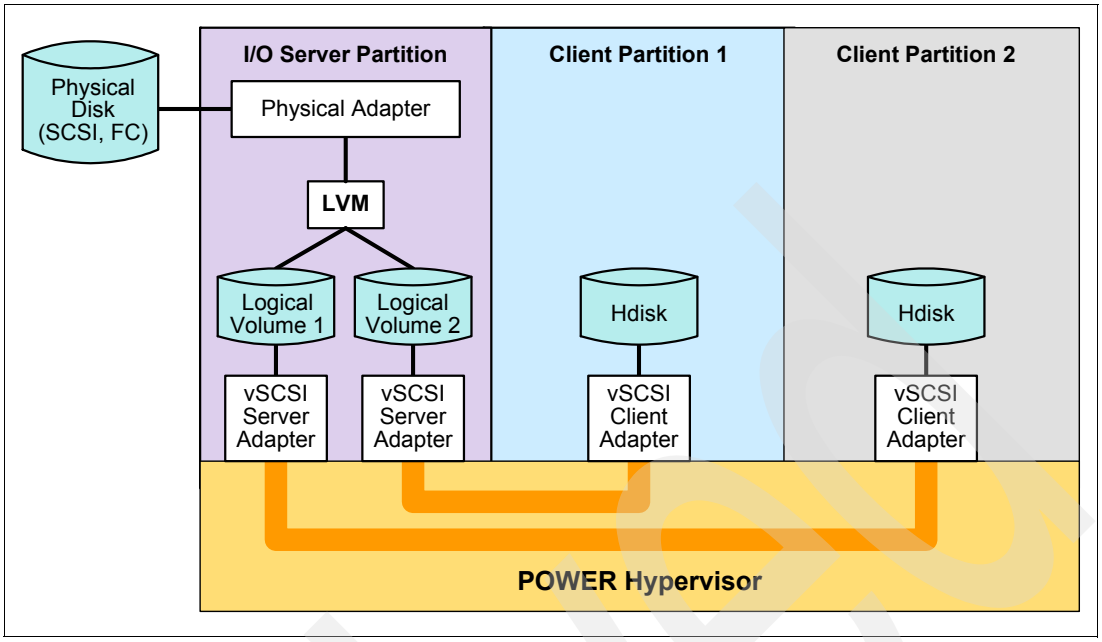


Figure 3-12 Architectural view of virtual SCSI

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 the specific storage devices supported for Virtual I/O Server, see:

`http://www14.software.ibm.com/webapp/set2/sas/f/vios/documentation/datasheet.html`

Virtual I/O Server functions

Virtual I/O Server includes a number of features, including monitoring solutions:

- Support for Live Partition Mobility on POWER6 and POWER7 processor-based systems with the PowerVM Enterprise Edition. More information about Live Partition Mobility can be found on 3.5.6, “PowerVM Live Partition Mobility” on page 97.
- Support for virtual SCSI devices backed by a file. These are then accessed as standard SCSI-compliant LUNs.
- Support for virtual Fibre Channel devices 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), SNMP v3 (Simple Network Management Protocol) and LDAP (Lightweight Directory Access Protocol client functionality).
- System Planning Tool (SPT) and Workload Estimator are designed to ease the deployment of a virtualized infrastructure. More information about the System Planning Tool is provided in 3.6, “System Planning Tool” on page 102.
- IBM Systems Director and a number of pre-installed IBM Tivoli® agents are included, such as Tivoli Identity Manager (TIM), in order to allow easy integration into an existing Tivoli Systems Management infrastructure, and Tivoli Application Dependency Discovery Manager (ADDM), which automatically creates and maintains application infrastructure maps, including dependencies, change histories, and deep configuration values.
- vSCSI eRAS is also available.

- Additional Command Line Interface (CLI) statistics in **svmon**, **vmstat**, **fcstat** and **topas**.
- Monitoring solutions to help manage and monitor the Virtual I/O Server and shared resources. New commands and views provide additional metrics for memory, paging, processes, Fibre Channel HBA statistics, and virtualization.

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.

3.5.5 PowerVM Lx86

PowerVM Lx86 supports the installation and running of most 32-bit x86 Linux applications on any POWER5, POWER5+, POWER6, POWER6+, and POWER7 processor-based servers, or IBM Power Architecture technology-based blade servers. PowerVM Lx86 creates a virtual x86 environment, within which the Linux on Intel® applications can run. Currently, a virtual PowerVM Lx86 environment supports SUSE Linux or Red Hat Linux x86 distributions. The translator and the virtual environment run strictly within user space. No modifications to the POWER kernel are required. PowerVM Lx86 does not run the x86 kernel on the POWER system and is not a virtual machine. Instead, x86 applications are encapsulated so that the operating environment appears to be Linux on x86, even though the underlying system is a Linux on POWER system.

PowerVM Lx86 release 1.4 supports all POWER7 processors-based servers.

Note: RHEL5 only supports POWER7 processor-based servers in POWER6 Compatibility Mode. Customers will need to use SLES11 to take full advantage of the POWER7 technology.

For more information about PowerVM Lx86, see the following website:

`http://www14.software.ibm.com/webapp/set2/sas/f/lx86/related/home.html`

3.5.6 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.

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 allows your end users to continue their work without disruption.
- Perform 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.
- Optimize server workloads:
  - Consolidation: You can consolidate workloads running on various small, under-utilized 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.

- Use Live Partition Mobility for a migration from POWER6 to POWER7 processor-based servers without any downtime of your applications.

Mobile partition operating system requirements

The operating system running in the mobile partition must be AIX or Linux. The Virtual I/O Server partition itself cannot be migrated. All versions of AIX and Linux supported on the IBM POWER7 processor-based servers also support partition mobility.

Source and destination system requirements

The source partition must be one that only has virtual devices. If there are any physical devices in its allocation, they must be removed before the validation or migration is initiated. An NPIV device is considered virtual and is compatible with partition migration.

The POWER Hypervisor must support the Partition Mobility functionality, also called migration process. POWER6 and POWER7 processor-based Hypervisors have this capability and firmware must be at firmware level eFW3.2 or later. All POWER7 processor-based Hypervisors support Partition Mobility. Source and destination systems can have various firmware levels, but they must be compatible with each other.

It is possible to migrate partitions back and forth between POWER6 and POWER7 processor-based servers. Partition Mobility leverages the POWER6 Compatibility Modes provided by POWER7 processor-based servers. On the POWER7 processor technology based server, the migrated partition is then executing in POWER6 or POWER6+ Compatibility Mode.

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, you can perform these steps:

1. You set the percolating preferred processor compatibility mode to the default mode. When you activate the logical partition on the POWER6 processor-based server, it runs in the POWER6 mode.
2. You 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. You 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 POWER hypervisor determines that the most fully featured mode supported by the operating environment installed in the logical partition is the POWER7 mode and changes the current mode of the logical partition to the POWER7 mode.

At this point, 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: For an exhaustive presentation of the supported migrations, see the “Migration combinations of processor compatibility modes for active Partition Mobility” web page:

http://publib.boulder.ibm.com/infocenter/powersys/v3r1m5/topic/p7hc3/iphc3pcmcombosact.htm

The Virtual I/O Server on the source system provides 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 POWER hypervisor; it is created and managed automatically by the HMC and will be configured on both the source and destination Virtual I/O Servers designated as the mover service partitions for the mobile partition to participate in active mobility. Other requirements include a similar Time of Day on each server, systems must not be running on battery power, shared storage (turn off SCSI reservation at any external hdisk using reserve_policy=no_reserve), and all logical partitions must be on the same open network with RMC established to the HMC.

The HMC is used to configure, validate, and orchestrate. You will use the HMC to configure the Virtual I/O Server as an MSP and to configure the VASI device. An HMC wizard validates your configuration and identifies issues that will cause the migration to fail. During the migration, the HMC controls all phases of the process.

Improved Live Partition Mobility benefits

The possibility to move partitions between POWER6 and POWER7 processor-based servers greatly facilitates the deployment of POWER7 processor technology based servers:

- Installation of the new server can be performed while the application is executing on the POWER6 server. When the POWER7 processor technology based server is ready, the application can be migrated to its new hosting server without application down time.
- When adding POWER7 processor technology based servers to a POWER6 environment, you get the additional flexibility to perform workload balancing across the whole set of POWER6 and POWER7 processor technology based servers.
- When performing server maintenance, you get the additional flexibility to utilize POWER6 Servers for hosting applications usually hosted on POWER7 processor technology based servers, and the opposite situation, allowing you to perform this maintenance with no application planned down time.

Note: A new Trial PowerVM Live Partition Mobility feature (#ELPM) is available on multiple POWER7 machine type models, which will enable clients to upgrade from PowerVM Standard Edition to PowerVM Enterprise Edition at no charge for 60 days. At the conclusion of the trial period clients can place an upgrade order for permanent PowerVM Enterprise Edition to maintain continuity. At the end of the trial period (60 days), the client's system will automatically return to PowerVM Standard Edition.

For more information about Live Partition Mobility and how to implement it, see *IBM PowerVM Live Partition Mobility*, SG24-7460.

3.5.7 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 only available with the PowerVM Enterprise edition.

The physical memory of a IBM Power System server can be assigned to multiple partitions either in a dedicated mode or a shared mode. The system administrator has the capability to assign part of the physical memory to a partition and other 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, it is the system administrator's task to optimize available memory distribution among partitions. When a partition suffers degradation due to memory constraints and other partitions have unused memory, the administrator can react manually by issuing a dynamic memory reconfiguration.

With a shared memory model, it is the system that 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 exploited 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 various operating systems. For example, AIX partitions can take advantage of Active Memory Expansion while other operating systems take advantage of Active Memory Sharing.

For additional information regarding Active Memory Sharing, see *PowerVM Virtualization Active Memory Sharing*, REDP-4470.

3.5.8 NPIV

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, which acts only as a pass-through managing the data transfer through the POWER Hypervisor.

Each partition using 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 additional information about NPIV, see the following Redbooks publications:

- *IBM PowerVM Virtualization Managing and Monitoring*, SG24-7590
- *PowerVM Migration from Physical to Virtual Storage*, SG24-7825

NPIV is supported with PowerVM Express, Standard, and Enterprise Editions, on all POWER7 processor-based servers and IBM Power Architecture technology-based blade servers, hosted by VIOS 2.1.3 or later and LPARS using AIX 5.3, AIX 6.1, IBM i 6.1.1, SLES 11, and RHEL 5.4, or later.

3.5.9 Operating System support for PowerVM

Table 3-6 summarizes the PowerVM features supported by the operating systems compatible with the POWER7 processor-based servers.

Table 3-6 PowerVM 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 5.5	SLES 10 SP3	SLES 11 SP1
Simultaneous Multi-Threading (SMT)	Yes[1]	Yes[2]	Yes	Yes[5]	Yes	Yes[1]	Yes[1]	Yes
DLPAR I/O adapter add/remove	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
DLPAR processor add/remove	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 5.5	SLES 10 SP3	SLES 11 SP1
DLPAR memory add	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
DLPAR memory remove	Yes	Yes	Yes	Yes	Yes	No	No	Yes
Capacity on Demand [3]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Micro-Partitioning	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Shared Dedicated Capacity	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Multiple Shared Processor Pools	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Virtual I/O Server	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
IVM	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Virtual SCSI	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Virtual Ethernet	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
NPIV	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
Live Partition Mobility	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Workload Partitions	No	Yes	Yes	No	No	No	No	No
Active Memory Sharing	Yes	Yes	Yes	Yes	Yes	No	No	Yes
Active Memory Expansion	No	Yes[4]	Yes	No	No	No	No	No

1 Only supports two threads.
2 AIX 6.1 up to TL4 SP2 only supports two threads, and supports four threads as of TL4 SP3.
3 Available on selected models.
4 On AIX 6.1 with TL4 SP2 and later.
5 IBM i 6.1.1 and up support SMT4.

3.5.10 POWER7-specific Linux programming support

The IBM Linux Technology Center (LTC) contributes to the development of Linux by providing support for IBM hardware in Linux distributions. In particular, the LTC makes tools and code available to the Linux communities to take advantage of the POWER7 technology, and develop POWER7 optimized software.

Table 3-7 summarizes the support of specific programming features for various versions of Linux.

Table 3-7 Linux support for POWER7 features

Features	Linux releases			Comments
	SLES 10 SP3	SLES 11 SP1	RHEL 5.5	
POWER6 compatibility mode	Yes	Yes	Yes	
POWER7 mode	No	Yes	No	
Strong Access Ordering	No	Yes	No	To improve Lx86 performance

Features	Linux releases			Comments
	SLES 10 SP3	SLES 11 SP1	RHEL 5.5	
Scale to 256 cores / 1024 threads	No	Yes	No	Base OS support available
4-way SMT	No	Yes	No	
VSX Support	No	Partial	No	
Distro toolchain mcpu/mtune=p7	No	No	No	
Advance Toolchain Support	Yes; execution restricted to POWER6 instructions	Yes	Yes; execution restricted to POWER6 instructions	Alternative IBM gnu Toolchain
64k base page size	No	Yes	Yes	
Tickless idle	No	Yes	No	Improved energy utilization and virtualization of partially to fully idle partitions

For information regarding Advanced Toolchain, see the “How to use Advance Toolchain for Linux” website:

http://www.ibm.com/developerworks/wikis/display/hpccentral/How+to+use+Advance+Toolchain+for+Linux+on+POWER

You can also see the University of Illinois Linux on Power Open Source Repository:

http://ppclinux.ncsa.illinois.edu
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

3.6 System Planning Tool

The IBM System Planning Tool (SPT) helps you design a system or systems to be partitioned with logical partitions. You can also plan for and design non-partitioned systems 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:

- To create reports
- As input to the IBM configuration tool (e-Config)
- To create and deploy partitions on your systems automatically

System plans generated by the SPT can be deployed on the system by the Hardware Management Console (HMC) or the Integrated Virtualization Manager (IVM).

Note: Ask your IBM representative or Business Partner representative 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 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 upon any of the following considerations:

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

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

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

Use the IBM System Planning Tool to estimate POWER Hypervisor requirements and determine the memory resources required for all partitioned and non-partitioned servers.

Figure 3-13 shows the estimated Hypervisor memory requirements based on sample partition requirements.

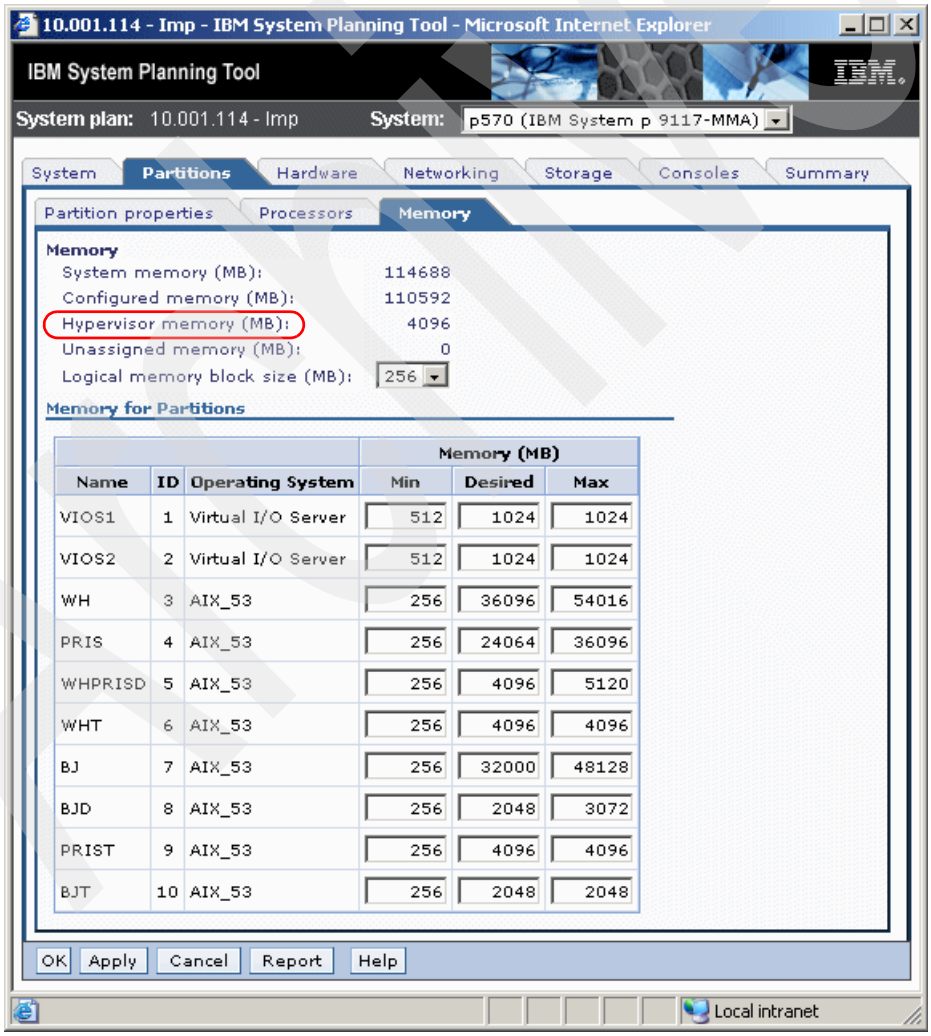


Figure 3-13 IBM System Planning Tool window showing Hypervisor memory requirements

For the SPT and its supporting documentation, see the IBM System Planning Tool site at:

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

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, provides the possibility to improve your architecture's total cost of ownership (TCO) by reducing unplanned down time.

The elements of RAS can be described as follows:

- Reliability: Indicates how infrequently a defect or fault in a server manifests itself.
- 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 users and services 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 a number of features designed to avoid or correct environmentally induced, single-bit, intermittent failures as well as handle solid faults in components, including selective redundancy to tolerate certain faults without requiring an outage or parts replacement.

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 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-fourth as many processor chips (and chip socket interfaces) as with a double CPU-per-processor design. Not only does this case reduce the total number of system components, it reduces the total amount of heat generated in the design, resulting in an additional reduction in required power and cooling components. 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; grade 3 defined as industry standard (off-the-shelf). As shown in Figure 4-1, using stringent design criteria and an extensive testing program, the IBM manufacturing team can produce grade 1 components that are expected to be ten 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.

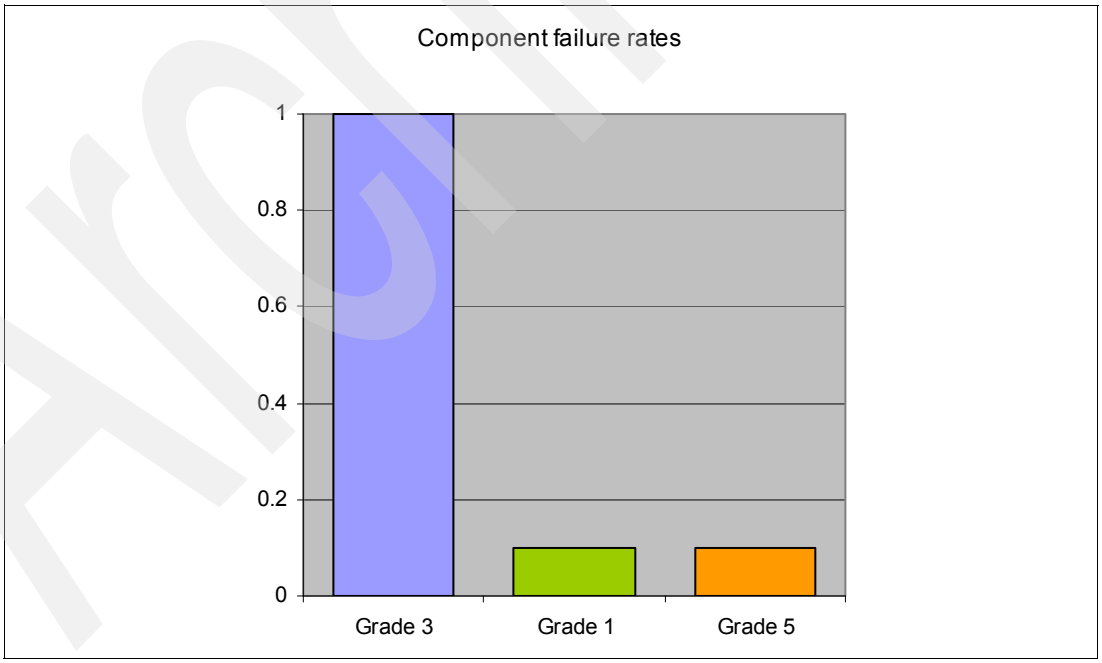


Figure 4-1 Component Failure rates

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; POWER processor-based systems are carefully packaged to ensure adequate cooling. Critical system components such as the POWER7 processor chips are positioned on printed circuit cards so they receive fresh air 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 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 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 in order to achieve maximum availability.

Tip: Check your configuration for optional redundant components before ordering your system.

4.2 Availability

IBM hardware and microcode capability to continuously monitor execution of hardware functions is generally described as the process of First Failure Data Capture (FFDC). This process includes the strategy of predictive failure analysis, which refers to the ability to track intermittent correctable errors and to vary components off-line before they reach the point of hard failure causing a system outage, and without the need to recreate the problem.

The POWER7 family of systems continues to offer and introduce significant enhancements designed to increase system availability to drive towards a high availability objective with hardware components that can perform the following automatic functions:

- Self-diagnose and self-correct during run time
- Automatically reconfigure to mitigate potential problems from suspect hardware
- Self-heal or automatically substitute good components for failing components

Note: POWER7 processor-based servers are independent of the operating system for error detection and fault isolation within the central electronics complex.

Throughout this chapter we describe IBM POWER7 processor-based systems technologies focused on keeping a system up and running. For a specific set of functions focused on detecting errors before they become serious enough to stop computing work, see 4.3.1, “Detecting” on page 117.

4.2.1 Partition availability priority

Also available is the ability to assign availability priorities to partitions. If an alternate processor recovery event requires spare processor resources and there are no other means of obtaining the spare resources, the system determines which partition has the lowest priority and attempts to claim the needed resource. On a properly configured POWER processor-based server, this approach allows that capacity to first be obtained from a low priority partition instead of a high priority partition.

This capability is relevant to the total system availability because it gives the system an additional stage before an unplanned outage. In the event that insufficient resources exist to maintain full system availability, these servers attempt to maintain partition availability by user-defined priority.

Partition availability priority is assign to partitions using a *weight value* or integer rating. The lowest priority partition is rated at 0 (zero) and the highest priority partition is valued at 255. The default value is set at 127 for standard partitions and 192 for Virtual I/O Server (VIOS) partitions. You can vary the priority of individual partitions.

Partition availability priorities can be set for both dedicated and shared processor partitions. The POWER Hypervisor uses the relative partition weight value among active partitions to favor higher priority partitions for processor sharing, adding and removing processor capacity, and favoring higher priority partitions for normal operation.

Note that the partition specifications for *minimum*, *desired*, and *maximum* capacity are also taken into account for capacity-on-demand options, and if total system-wide processor capacity becomes disabled because of deconfigured failed processor cores. For example, if total system-wide processor capacity is sufficient to run all partitions, at least with the minimum capacity, the partitions are allowed to start or continue running. If processor capacity is insufficient to run a partition at its minimum value, then starting that partition results in an error condition that must be resolved.

4.2.2 General detection and deallocation of failing components

Runtime correctable or recoverable errors are monitored to determine if there is a pattern of errors. If these components reach a predefined error limit, the service processor initiates an action to deconfigure the faulty hardware, helping to avoid a potential system outage and to enhance system availability.

Persistent deallocation

To enhance system availability, a component that is identified for deallocation or deconfiguration on a 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, runtime unrecoverable hardware faults can be deconfigured from the system after the first occurrence. The system can be rebooted immediately after failure and resume operation on the remaining stable hardware. This way prevents the same faulty hardware from affecting system operation again; the repair action is deferred to a more convenient, less critical time.

The following persistent deallocation functions are included:

- Processor
- L2/L3 cache lines (cache lines are dynamically deleted)
- Memory
- Deconfigure or bypass failing I/O adapters

> **Note:** The auto-restart (reboot) option has to be enabled from the Advanced System Manager Interface or from the Control (Operator) Panel. Figure 4-2 shows this option using the ASMI.

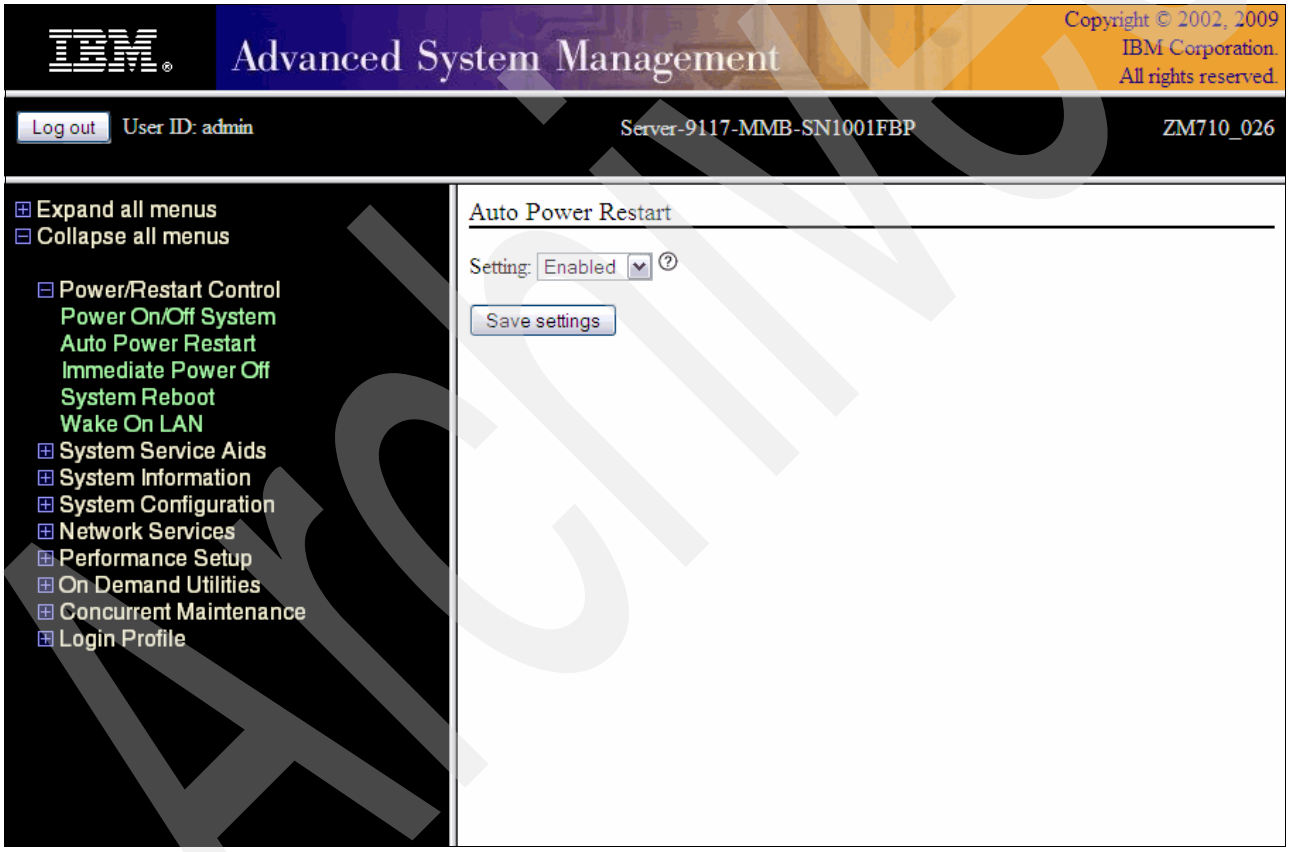


Figure 4-2 AMSI Auto Power Restart setting window interface

Processor instruction retry

As in POWER6, the POWER7 processor has the ability to do processor instruction retry and alternate processor recovery for a number of core related faults. Doing this significantly reduces exposure to both permanent and intermittent errors in the processor core.

Intermittent errors, often due to cosmic rays or other sources of radiation, are generally not repeatable.

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.

Note that on IBM systems prior to POWER6, such an error typically caused a checkstop.

Alternate processor retry

Hard failures are more difficult, being permanent errors that will be replicated each time the instruction is repeated. Retrying the instruction does not help in this situation because the instruction will continue to fail.

As in POWER6, POWER7 processors have the ability to extract the failing instruction from the faulty core and retry it elsewhere in the system for a number of faults, after which the failing core is 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. Then the POWER Hypervisor dynamically deconfigures the failing core and is called out for replacement. 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 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 full checkstop goes into effect.

Even with all these availability enhancements to prevent processor errors from affecting system-wide availability into play, there will be errors 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 very inadequate for protecting the much larger system main store. Therefore, a variety of protection methods are used in POWER processor-based systems to avoid uncorrectable errors in memory.

Memory protection plans must take into account many factors, including these:

- Size
- Desired performance
- Memory array manufacturing characteristics

POWER7 processor-based systems have a number of 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 9 DRAM chips on the Power 720 and 740). 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/Single Error Correction ECC implementation found on the POWER6 processor-based systems).

- Hardware scrubbing:

  Hardware scrubbing is a method used to deal with 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.

- 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 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 which is determined to be faulty.

Figure 4-3 shows a POWER7 processor chip, with its memory interface, consisting of two controllers and four DIMMs per controller. Advanced memory buffer chips are exclusive to IBM and help to increase performance, acting as read/write buffers. On the Power 770 and 780, the advanced memory buffer chips are integrated to the DIMM that they support. Power 750 and 755 uses only one memory controller, Advanced memory buffer chips are on the system planar and support two DIMMs each.

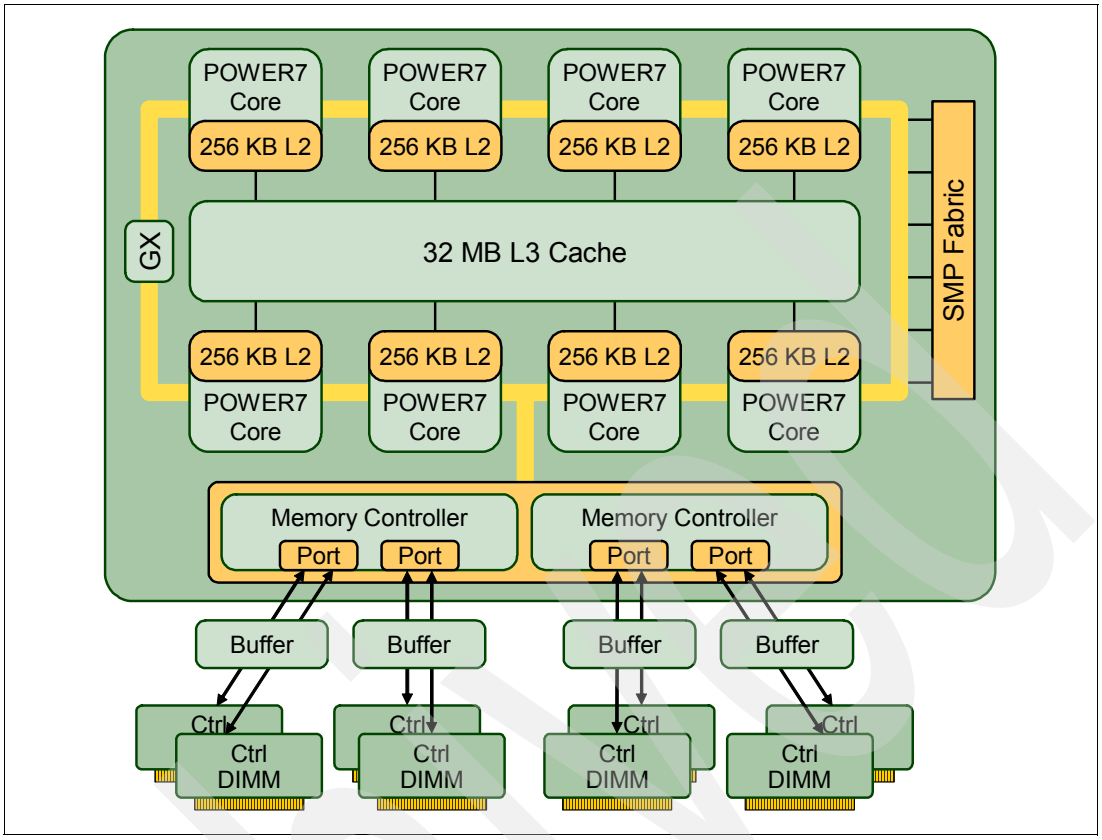


Figure 4-3 POWER7 memory subsystem

Memory page deallocation

While coincident cell errors in separate memory chips are statistically rare, IBM POWER7 processor-based systems can contain these errors using a memory page deallocation scheme for partitions running IBM AIX and the IBM i operating systems as well as for memory pages owned by the POWER Hypervisor. If a memory address experiences an uncorrectable or repeated correctable single cell error, the Service Processor sends the memory page address to the POWER Hypervisor to be marked for deallocation.

Pages 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 page(s) 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 end users and applications.

The POWER Hypervisor maintains a list of pages marked for deallocation during the current platform IPL. During a partition IPL, the partition receives a list of all the bad pages in its address space. In addition, if memory is dynamically added to a partition (through a dynamic LPAR operation), the POWER Hypervisor warns the operating system when memory pages are included that need to be deallocated.

Finally, If an uncorrectable error in memory is discovered, the logical memory block associated with the address with the uncorrectable error is marked for deallocation by the POWER Hypervisor. This deallocation will take effect 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 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 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 it is not to be 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 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 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, as well as 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 earlier. 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 perform 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 subsequently written back to the L2 and L3 cache.

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 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

While it is rare, an uncorrectable data error can occur in memory or a 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, for example:

- Data in the instruction L1 cache is never modified within the cache itself. Therefore, an uncorrectable error discovered in the cache is treated like an ordinary cache miss, and correct data is loaded from the L2 cache.
- 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 called Special Uncorrectable Error (SUE) handling is used to prevent an uncorrectable error in memory or cache from immediately causing the system to terminate. Rather, the system tags the data and determines whether it will ever be used again:

- If the error is irrelevant, it will not force a checkstop.
- If the data is used, termination can be limited to the program / kernel or hypervisor owning the data; or freeze of the I/O adapters 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 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.

It is only in the case where the corrupt data is used by the POWER Hypervisor that the entire system must be rebooted, thereby preserving overall system integrity.

Depending on system configuration and source of the data, errors encountered during I/O operations 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 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 extended error handling

IBM estimates that PCI adapters can account for a significant portion of the hardware-based errors on a large server. Whereas 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 extended error handling (EEH) enabled adapters respond to a special data packet generated from the affected PCI slot hardware by calling system firmware, which will examine the affected bus, allow the device driver to reset it, and continue without a system reboot. 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, while for PCI-X use ECC.

Figure 4-4 shows the location and various mechanisms used throughout the I/O subsystem for PCI extended error handling.

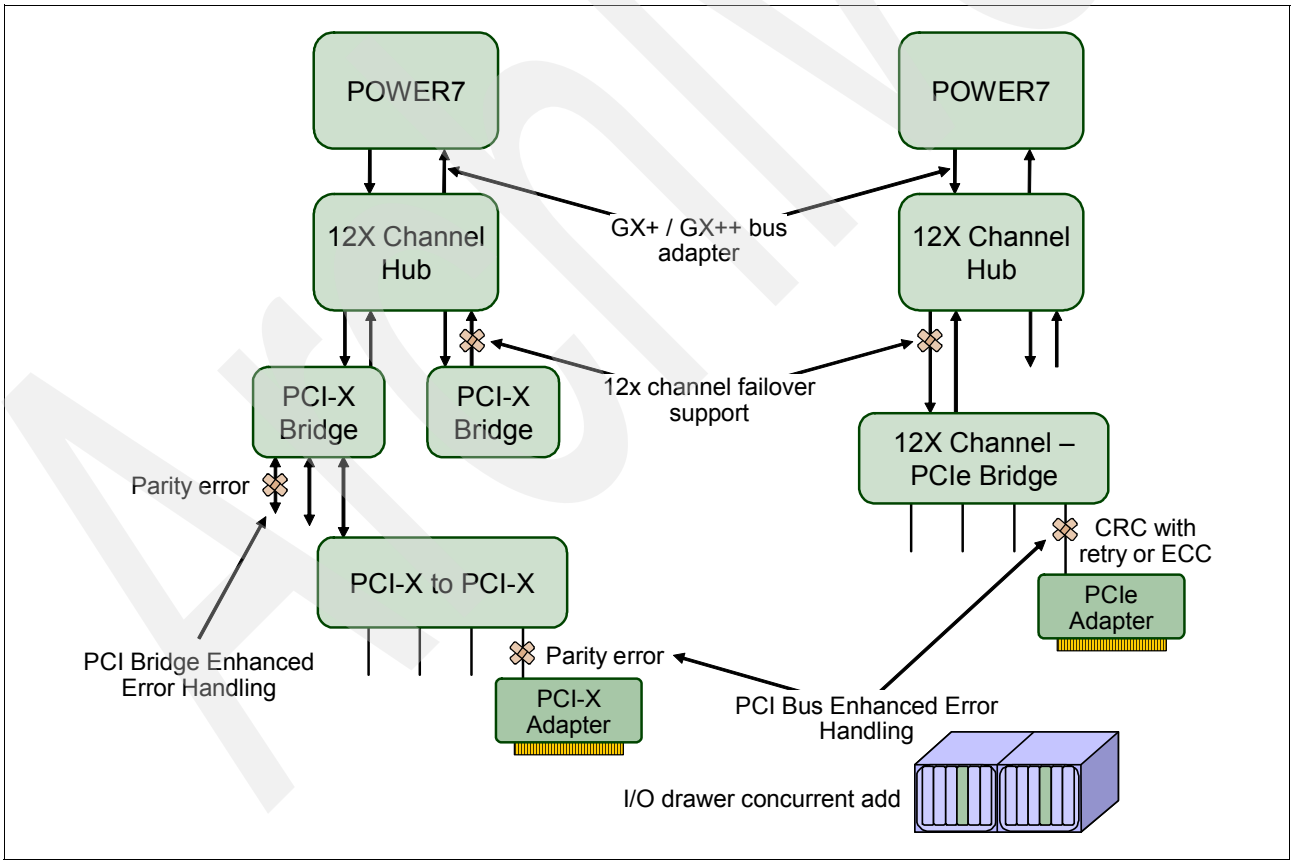


Figure 4-4 PCI extended error handling

4.3 Serviceability

IBM Power Systems design considers both IBM and client needs. The IBM Serviceability Team has enhanced the base service capabilities and continues to implement a strategy that incorporates best-of-breed service characteristics from diverse IBM systems offerings.

Serviceability includes system installation, system upgrades and downgrades (MES), and system maintenance and repair.

The goal of the IBM Serviceability Team is to design and provide the most efficient system service environment, which includes the following benefits:

- Easy access to service components; design for Customer Set Up (CSU), Customer Installed Features (CIF), and Customer Replaceable Units (CRU)
- On demand service education
- Error detection and fault isolation (ED/FI)
- First-failure data capture (FFDC)
- An automated guided repair strategy that uses common service interfaces for a converged service approach across multiple IBM server platforms

By delivering on these goals, 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 that are found in all POWER processor-based systems.

The term *servicer*, when used in the context of this document, denotes the person tasked with performing service related actions on a system. For an item designated as a Customer Replaceable Unit (CRU), the servicer might be the client. In other cases, for Field Replaceable Unit (FRU) items, the servicer might be an IBM representative or an authorized warranty service provider.

Service can be divided into three main categories:

Service Components	The basic service related building blocks
Service Functions	Service procedures or processes containing one or more service components
Service Operating Environment	The specific system operating environment which specifies how service functions are provided by the various service components

The basic component of service is a *Serviceable Event*.

Serviceable events are platform, regional, and local error occurrences that require a service action (repair). This action can include a *call home* to report the problem so that the repair can be assessed by a trained service representative. In all cases, the client is notified of the event. Event notification includes a clear indication of when servicer intervention is required to rectify the problem. The intervention might be a service action that the client can perform or it might require a service provider.

Serviceable events are classified as follows:

1. Recoverable: This is a correctable resource or function failure. The server remains available, but there might be a decrease in operational performance available for the client's workload (applications).
2. Unrecoverable: This is an uncorrectable resource or function failure. In this instance, there is potential degradation in availability and performance, or loss of function to the client's workload.
3. Predictable (using thresholds in support of Predictive Failure Analysis): This is a determination that continued recovery of a resource or function might lead to degradation of performance or failure of the client's workload. Although the server remains fully available, if the condition is not corrected, an unrecoverable error might occur.
4. Informational: This is notification that a resource or function:
   - Is *out-of* or *returned-to* specification and might require user intervention.
   - Requires user intervention to complete a system task(s).

Platform errors are faults that affect all partitions in various ways. They are detected in the CEC by the Service Processor, the System Power Control Network, or the Power Hypervisor. When a failure occurs in these components, the POWER Hypervisor notifies each partition's operating system to execute any required precautionary actions or recovery methods. The OS is required to report these kinds of errors as serviceable events to the Service Focal Point application because, by the definition, they affect the partition.

Platform errors are faults related to the following components:

- The Central Electronics Complex (CEC); that part of the server composed of the central processor units, memory, storage controls, and the I/O hubs.
- The power and cooling subsystems.
- The firmware used to initialize the system and diagnose errors.

Regional errors are faults that affect some, but not all partitions. They are detected by the POWER Hypervisor or the Service Processor. Examples of these are RIO bus, RIO bus adapter, PHB, multi-adapter bridges, I/O hub, and errors on I/O units (except adapters, devices, and their connecting hardware).

Local errors are faults detected in a partition (by the partition firmware or the operating system) for resources owned only by that partition. The POWER Hypervisor and Service Processor are not aware of these errors. Local errors might include "secondary effects" that result from platform errors preventing partitions from accessing partition-owned resources. Examples include PCI adapters or devices assigned to a single partition. If a failure occurs to one of these resources, only a single operating system partition need be informed.

This section provides an overview of the progressive steps of error detection, analysis, reporting, notifying, and repairing that are 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 does not have the opportunity to evaluate and act if necessary. POWER processor-based systems employ IBM System z® server-inspired error detection mechanisms that extend from processor cores and memory to power supplies and hard drives.

Service processor

The service processor (or also referred as flexible service processor) is a microprocessor, which is powered separately from the main instruction processing complex. The service processor provides the capabilities for the following features:

- POWER Hypervisor (system firmware) and Hardware Management Console connection surveillance
- Various 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. 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 under 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 when:

  – 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.

- 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 POWER Hypervisor firmware during the boot process and watches for loss of control during system operation. It also allows the POWER Hypervisor to monitor service processor activity. The service processor can take appropriate action, including calling for service, when it detects the POWER Hypervisor firmware has lost control. Likewise, the POWER Hypervisor can request a service processor repair action if necessary.

- Availability:

  The auto-restart (reboot) option, when enabled, can reboot the system automatically following an unrecoverable firmware error, firmware hang, hardware failure, or environmentally induced (AC power) failure.

- Fault monitoring:

  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 Advanced System Management Interface (ASMI):

  Concurrent access allows nondisruptive abilities to change system default parameters, interrogate service processor progress and error logs, set and reset server indicators, (Guiding Light for midrange and high-end servers, Light Path for low end servers), indeed, accessing all service processor functions without having to power-down the system to the standby state. This way 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 subsystems.

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 run time 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, run-time 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

The first-failure data capture (FFDC) error isolation technique 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 (FIRs) 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 the server logic and data paths to ensure that potential errors can be quickly identified and accurately tracked to a field-replaceable unit (FRU).

This proactive diagnostic strategy is a significant improvement over the classic, less accurate *reboot and diagnose* service approaches.

Figure 4-5 shows a schematic of a fault isolation register implementation.

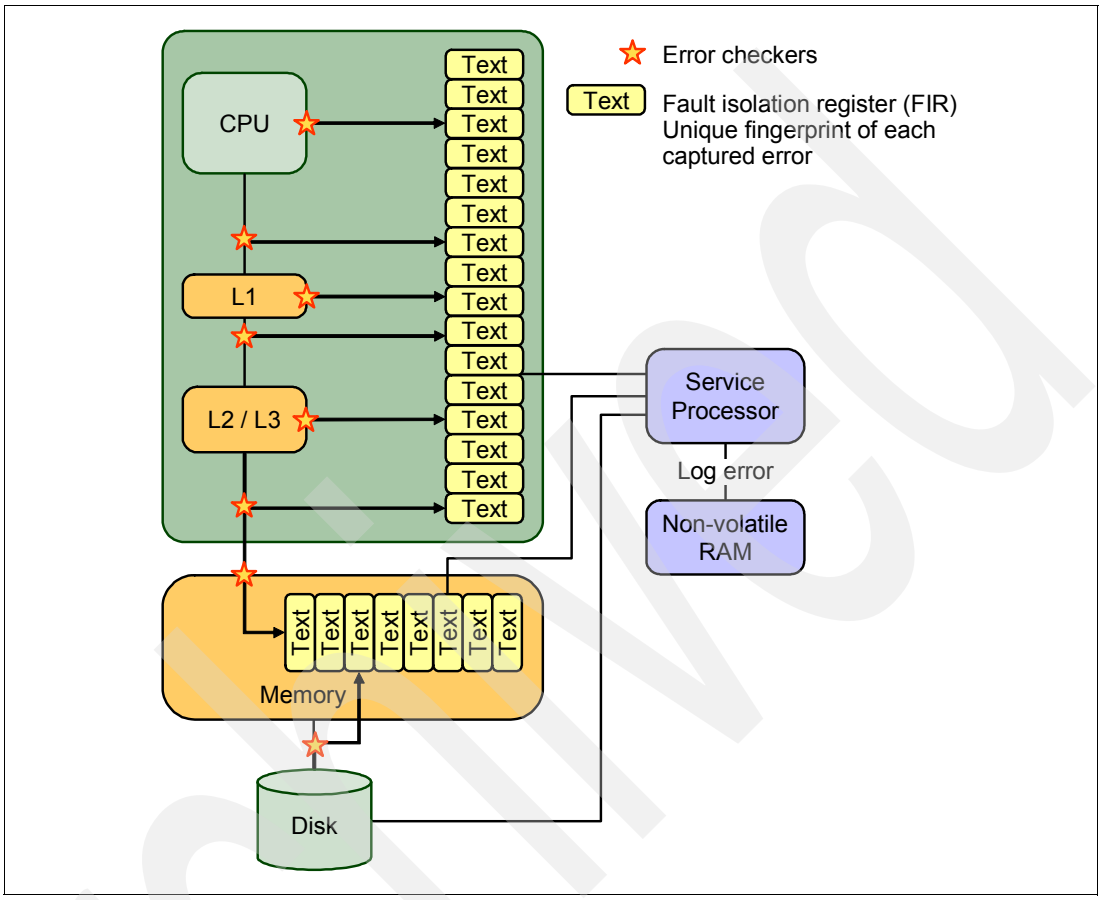


Figure 4-5 Schematic of FIR implementation

Note: First Failure Data Capture plays a critical role in delivering servers that can self-diagnose and self-heal. Using thousands of checkers (diagnostic probes) deployed at critical junctures throughout the server, the system effectively “traps” hardware errors at system run time.

The separately powered Service Processor is then used to analyze the checkers and perform problem determination. Using this method, IBM no longer has to rely on an intermittent “reboot and retry” error detection strategy, but knows with a degree of certainty which part is having problems. In this automated approach, 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 FRU (part) call, and manual interpretation of the data is not normally required.

This architecture is also the basis for the IBM predictive failure analysis, because the Service Processor can now count, and log, intermittent component errors, and can deallocate or take other corrective actions when an error threshold is reached.

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) in order to determine the root cause of the error event.

Root cause analysis might indicate that the event is recoverable, meaning that a service action point or need for repair has not been reached. Alternatively, it might indicate that a service action point has been reached, where the event exceeded a pre-determined threshold or was unrecoverable. Based on the isolation analysis, recoverable error threshold counts might 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 has been reached, a request for service is initiated through an error logging component.

4.3.2 Diagnosing

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.

Boot time

When an IBM Power Systems server powers up, the service processor initializes system hardware. Boot-time diagnostic testing uses a multi-tier approach for system validation, starting with managed low level diagnostics that are supplemented with system firmware initialization and configuration of I/O hardware, followed by OS-initiated software test routines. The following boot-time diagnostic routines are included:

- 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 in order to ensure correct operation, based on the way the system was powered off, or on the boot-time selection menu.

Run time

All 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 have been developed and improved over many generations of POWER processor-based servers, and enable quick and accurate predefined responses to both actual and potential system problems.

The service processor correlates and processes runtime error information, using logic derived from IBM engineering expertise to count recoverable errors (called thresholding) and predict when corrective actions must be automatically initiated by the system. The following actions are included:

- 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 the 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 a number of 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 has the ability to automatically send out 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 an HMC is attached, an ELA routine analyzes the error, forwards the event to the Service Focal Point (SFP) application running on the HMC, and notifies 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 an IBM service organization. Client contact information and specific system-related data such as the machine type, model, and serial number, along with error log data related to the failure are sent to IBM Service.

Error logging and analysis

When the root cause of an error has been identified by a fault isolation component, an error log entry is created with basic data such as the following examples:

- An error code uniquely describing the error event
- The location of the failing component
- The part number of the component to be replaced, including pertinent data such as engineering and manufacturing levels
- Return codes
- Resource identifiers
- FFDC data

Data containing information about the effect that the repair will have on the system is also included. Error log routines in the operating system can then use this information and decide to call home to contact service and support, send a notification message, or continue without an alert.

Remote support

The Remote Management and Control (RMC) application is delivered as part of the base operating system, including the operating system running on the Hardware Management Console. RMC provides a secure transport mechanism across the LAN interface between the operating system and the Hardware Management Console and is used by the operating system diagnostic application for transmitting error information. It performs a number of other functions also, but these are not used for the service infrastructure.

Service Focal Point

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 must only be reported once, regardless of how many logical partitions experience the potential effect of the error. The Manage Serviceable Events task on the Hardware Management Console (HMC) is responsible for aggregating duplicate error reports, and ensures that all errors are recorded for review and management.

When a local or globally reported service request is made to the operating system, the operating system diagnostic subsystem uses the Remote Management and Control Subsystem (RMC) to relay error information to the Hardware Management Console. For global events (platform unrecoverable errors, for example) the service processor will also forward error notification of these events to the Hardware Management Console, providing a redundant error-reporting path in case of errors in the RMC network.

The first occurrence of each failure type is recorded in the Manage Serviceable Events task on the Hardware 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 collected is dependent 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 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 HMC upon the reboot. Specific HMC information is included as part of the information that can optionally be sent to IBM support for analysis. If additional information relating to the dump is required, or if it becomes necessary to view the dump remotely, the HMC dump record notifies the IBM support center regarding on which HMC the dump is located.

4.3.4 Notifying

After a Power Systems server has detected, diagnosed, and reported an error to an appropriate aggregation point, it then takes steps to notify the client, and if necessary, the IBM support organization. Depending upon the assessed severity of the error and support agreement, this 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, for example:

- 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 has happened that requires client awareness in the event the client wants to take further action. These events can always be reported back to IBM at the client's discretion.

Call home

A correctly configured POWER processor-based system can initiate an automatic or manual call from a client location to the IBM service and support organization with error data, server status, or other service-related information. The call-home feature invokes the service organization in order for the appropriate service action to begin, automatically opening a problem report, and in certain cases, also dispatching field support. This automated reporting provides faster and potentially more accurate transmittal of error information. Although configuring call-home is optional, clients are strongly encouraged to configure this feature in order to obtain the full value of IBM service enhancements.

Vital product data 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, manufacturing level of the parts, and so on. These records provide valuable information that can be used by remote support and service representatives, enabling them to provide assistance in keeping the firmware and software on the server up-to-date.

IBM problem management database

At the IBM support center, historical problem data is entered into the IBM Service and Support Problem Management database. All of the information that is related to the error, along with any service actions taken by the service representative, are recorded for problem management by the support and development organizations. The problem is then tracked and monitored until the system fault is repaired.

4.3.5 Locating and 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):
  - Terracotta-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 simple tools such as flathead screw drivers to service the hardware components.
- Positive retention: Positive retention mechanisms help to assure 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 run the risk of 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 supported by the Power 710, 720, 730, 740, 750, and the 755 servers. In the Light Path LED implementation, when a fault condition is detected on the 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 turning on 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 (staying on solid) the system fault LED, enclosure fault LED, and the component FRU fault LED.

After the repair, the LEDs shut off automatically if the problem is fixed.

Guiding Light

The enclosure and system identify LEDs that turn on solid and that can be used to follow the path from the system to the enclosure and down to the specific FRU.

Guiding Light uses a series of flashing LEDs, allowing a service provider to quickly and easily identify the location of system components. Guiding Light can also handle multiple error conditions simultaneously, which might be necessary in certain very complex high-end configurations.

In these situations, Guiding Light waits for the servicer's indication of what failure to attend first and then illuminates the LEDs to the failing component.

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 extend 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 them in performing maintenance actions. Service labels are found in various formats and positions, and are intended to transmit readily available information to the servicer during the repair process.

Various service labels and the purpose of each are described in the following list:

- Location diagrams are strategically located 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, CPUs, processor books, 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 spots 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 certain 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 710 and Power 730 is a four-row by 16-element LCD display 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 various buttons, enabling a service support representative or client to change various boot-time options and other limited service functions.

Concurrent maintenance

The IBM 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 have been 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 times like these, flexibility is crucial. The IBM 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, media devices, I/O drawers, GX adapter, and the operator panel. In addition, they support concurrent firmware fixpack updates when possible. The determination of whether a firmware fixpack release can be updated concurrently is identified in the *readme* file that is released with the firmware.

Firmware updates

Firmware updates for Power Systems are released in a cumulative sequential fix format, packaged as an RPM for concurrent application and activation. Administrators can install and activate many firmware patches without cycling power or rebooting the server.

When an HMC is connected to the system, the new firmware image is loaded using any of the following methods:

- IBM distributed media, such as a CD-ROM
- A Problem Fix distribution from the IBM Service and Support repository
- FTP from another server
- Download from the IBM Fix Central website:

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

IBM supports multiple firmware releases in the field. Therefore, under expected circumstances, a server can operate on an existing firmware release, using concurrent firmware fixes to stay up-to-date with the current patch level. Because changes to various server functions (for example, changing initialization values for chip controls) cannot occur during system operation, a patch in this area requires a system reboot for activation. Under normal operating conditions, IBM provides patches for an individual firmware release level for up to two years after first making the release code generally available. After this period, clients must plan to update in order to stay on a supported firmware release.

Activation of new firmware functions, as opposed to patches, require installation of a new firmware release level. This process is disruptive to server operations because it requires a scheduled outage and full server reboot.

In addition to concurrent and disruptive firmware updates, IBM also offers concurrent patches, which include functions that are not activated until a subsequent server reboot. A server with these patches can operate normally. The additional concurrent fixes can be installed and activated when the system reboots after the next scheduled outage.

Additional capability is being added to the firmware to be able to view the status of a system power control network background firmware update. This subsystem can update as necessary, as migrated nodes or I/O drawers are added to the configuration. The new firmware provides an interface to be able to view the progress of the update, and also control starting and stopping of the background update if a more convenient time becomes available.

Repair and verify

Repair and verify (R&V) is a system used to guide a service provider step-by-step through the process of repairing a system and verifying that the problem has been repaired. The steps are customized in the appropriate sequence for the particular repair for the specific system being repaired. Repair scenarios covered by repair and verify include:

- Replacing a defective field-replaceable unit (FRU)
- Reattaching a loose or disconnected component
- Correcting a configuration error
- Removing or replacing an incompatible FRU
- Updating firmware, device drivers, operating systems, middleware components, and IBM applications after replacing a part
- Installing a new part

Repair and verify procedures can be used by both 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.

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; a CD-ROM version is also available.

4.4 Manageability

Various functions and tools help manageability, and enable you to 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 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 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:

- Light Path and Guiding Light

  For more information, see “Light Path” on page 125 and “Guiding Light” on page 125.
- 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 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

Functions

The service processor is used to monitor and manage the system hardware resources and devices. The service processor checks the system for errors, ensuring the connection to the HMC for manageability purposes and accepting Advanced System Management Interface (ASMI) Secure Sockets Layer (SSL) network connections. The service processor provides the ability to view and manage the machine-wide settings by using the ASMI, and enables complete system and partition management from the HMC.

Note: The service processor enables a system that does not boot to be analyzed. The error log analysis can be performed from either the ASMI or the HMC.

The service processor uses two Ethernet 10/100Mbps ports. Note the following information:

- Both Ethernet ports are only visible to the service processor and can be used to attach the server to an HMC or to access the ASMI. The ASMI options can be accessed through an HTTP server that is integrated into the service processor operating environment.
- Both Ethernet ports 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 or HMC1 port is configured as 169.254.2.147
  - Service processor Eth1 or 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, PN, 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 (ASMI)

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 HMC. 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. 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 steps:

- Access the ASMI by using an HMC.

  If configured to do so, the HMC connects directly to the ASMI for a selected system from this task.

  To connect to the Advanced System Management interface from an HMC, follow these steps:

  a. Open Systems Management from the navigation pane.

  b. From the work pane, select one or more managed systems to work with.

  c. From the System Management tasks list, select **Operations Advanced System Management** (ASM).

- Access the ASMI by using a web browser.

  The web interface to the ASMI is accessible through Microsoft® Internet Explorer 6.0, Microsoft Internet Explorer 7, Netscape 7.1, Mozilla Firefox, or Opera 7.23 running on a PC or mobile computer that is connected to the service processor. The web interface is available during all phases of system operation, including the initial program load (IPL) and run time.

  However, a few of the menu options in the web interface are unavailable during IPL or run time to prevent usage or ownership conflicts if the system resources are in use during that phase. The ASMI provides a Secure Sockets Layer (SSL) web connection to the service processor. To establish an SSL connection, open your browser using the address:

  `https://<ip_address_of_service_processor>`

  (where <ip_address_of_service_processor> is the address of the service processor of your Power Systems server, such as 9.166.196.7).

  Note: To make the connection through Internet Explorer, click **Tools Internet Options**. Clear the **Use TLS 1.0** check box, and click **OK**.

- Access the ASMI using 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 standby state. The ASMI on an ASCII console is not available during various 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 view the LCD display. Figure 4-6 shows that the operator panel on a Power 710 and 730 is pulled out.

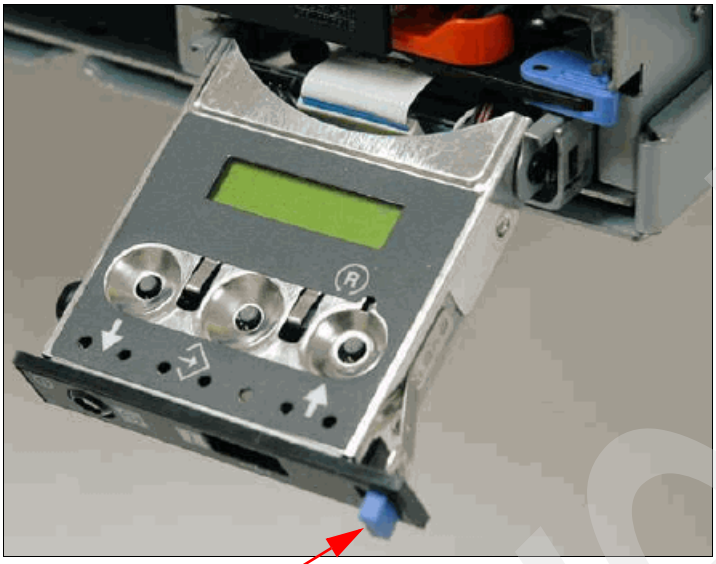


Figure 4-6 Operator panel is pulled out from the chassis

The operator panel includes features such as these:

- 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 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 available modes are as follows:

- Service mode:

  Requires a service mode boot of the system; enables the checking of system devices and features. Service mode provides the most complete checkout of the system resources. All system resources, except the SCSI adapter and the disk drives used for paging, can be tested.

- Concurrent mode:

  Enables the normal system functions to continue while selected resources are being checked. Because the system is running in normal operation, certain devices might require additional actions by the user or diagnostic application before testing can be done.

- Maintenance mode:

  Enables the checking of most system resources. Maintenance mode provides the same test coverage as service mode. The difference between the two modes is the way they are invoked. Maintenance mode requires that all activity on the operating system be stopped. The **shutdown -m** command is used to stop all activity on the operating system and put the operating system into maintenance mode.

The System Management Services (SMS) error log is accessible on the SMS menus. This error log contains errors that are found by partition firmware when the system or partition is booting.

You can access the service processor's error log on the ASMI menus.

You can also access the system diagnostics from an AIX Network Installation Management (NIM) server.

Note: When your order a Power Systems server, 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 leaves all applications up and running, using the IBM i **Start Service Tools** (STRSST) command (when signed onto IBM i with the appropriately secured user ID).

With DST and SST, you can look at various logs, run various diagnostics, or take various kinds of system dumps or other options.

Depending on the operating system, the service-level functions that you typically see when using the operating system service menus are as follows:

- 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 HMC 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 HMC, without determining whether other logical partitions also detect and report the errors. For example, if one logical partition reports an error for a shared resource, such as a managed system power supply, other active logical partitions might report the same error.

By using the Manage Serviceable Events task in the HMC, you can avoid long lists of repetitive call-home information by recognizing that these are repeated errors and consolidating them into one error.

In addition, you can use the Manage Serviceable Events task to initiate service functions on systems and logical partitions, including the exchanging of parts, configuring connectivity, and managing dumps.

The following functions are available through the Service Focal Point on the Hardware Management Console:

- Service Focal Point:
  - Managing serviceable events and service data
  - Managing service indicators
- Error Information:
  - OS Diagnostic
  - Service Processor
  - Service Focal Point
- LED Management menu
- Serviceable events analysis
- Repair and Verify:
  - Concurrent Maintenance
  - Deferred Maintenance
  - Immediate Maintenance
- Hot-Node Add, Hot-Node Repair, and Memory Upgrade
- FRU Replacement

- Managing firmware levels:
  - HMC
  - Server
  - Adapter
  - Concurrent Firmware updates
- Call Home/Customer Notification
- Virtualization
- I/O Topology view
- Generate dump
- Remote support (full access)
- Virtual operator panel

Service Focal Point Lite on the Integrated Virtualization Manager

The following functions are available through the Service Focal Point Lite on the Integrated Virtualization Manager:

- Service Focal Point-Lite:
  - Managing serviceable events and service data
  - Managing service indicators
- Call Home/Customer Notification (both not available yet)
- Error Information menu:
  - OS Diagnostic
  - Service Focal Point lite
- LED Management menu
- Managing firmware levels:
  - Server
  - Adapter
- Virtualization
- Generate dump (limited capability)
- Remote support (limited access)

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 one on the rear of a Power 710 and 730, if that system is not managed by an HMC.

Power Systems has a permanent firmware boot side, or A side, and a temporary firmware boot side, or B side. New levels of firmware must be installed on the temporary side first in order to test the update's compatibility with existing applications. When the new level of firmware has been 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 4-7 shows an example.

Figure 4-7 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 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 has been 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 & 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 fixpack		
DDD	Last disruptive fixpack		

The following example uses the convention:

```
01AL710_086 = POWER7 Entry Systems Firmware for 8233-E8B and 8236-E8B
```

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 Electronic Services and Electronic Service Agent

IBM has 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 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 that are traditionally used to access IBM Internet services and support. The news page enables you to gain easier access to IBM resources for assistance in resolving technical problems.

- IBM 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:

https://www.ibm.com/support/electronic/portal

4.5 Operating system support for RAS features

Table 4-2 provides an overview of a number of features for continuous availability that are supported by the various operating systems running on the POWER7 processor-based systems.

Table 4-2 Operating system support for RAS features

RAS feature	AIX 5.3	AIX 6.1	AIX 7.1	IBM i 6.1.1	IBM i 7.1	RHEL 5	SLES 10 SP 3	SLES 11 SP 1
System deallocation of failing components								
Dynamic Processor Deallocation	✓	✓	✓	✓	✓	✓	✓	✓
Dynamic Processor Sparing	✓	✓	✓	✓	✓	✓	✓	✓
Processor Instruction Retry	✓	✓	✓	✓	✓	✓	✓	✓
Alternate Processor Recovery	✓	✓	✓	✓	✓	✓	✓	✓
Partition Contained Checkstop	✓	✓	✓	✓	✓	✓	✓	✓
Persistent processor deallocation	✓	✓	✓	✓	✓	✓	✓	✓
GX++ bus persistent deallocation	✓	✓	✓	✓	✓	-	-	✓
PCI bus extended error detection	✓	✓	✓	✓	✓	✓	✓	✓
PCI bus extended error recovery	✓	✓	✓	✓	✓	Most	Most	Most
PCI-PCI bridge extended error handling	✓	✓	✓	✓	✓	-	-	-
Redundant RIO or 12x Channel link	✓	✓	✓	✓	✓	✓	✓	✓
PCI card hot-swap	✓	✓	✓	✓	✓	✓	✓	✓
Dynamic SP failover at run-time	✓	✓	✓	✓	✓	✓	✓	✓
Memory sparing with CoD at IPL time	✓	✓	✓	✓	✓	✓	✓	✓
Clock failover runtime or IPL	✓	✓	✓	✓	✓	✓	✓	✓
Memory availability								
64-byte ECC code	✓	✓	✓	✓	✓	✓	✓	✓
Hardware scrubbing	✓	✓	✓	✓	✓	✓	✓	✓
CRC	✓	✓	✓	✓	✓	✓	✓	✓
Chipkill	✓	✓	✓	✓	✓	✓	✓	✓
L1 instruction and data array protection	✓	✓	✓	✓	✓	✓	✓	✓
L2/L3 ECC & cache line delete	✓	✓	✓	✓	✓	✓	✓	✓
Special uncorrectable error handling	✓	✓	✓	✓	✓	✓	✓	✓
Fault detection and isolation								
Platform FFDC diagnostics	✓	✓	✓	✓	✓	✓	✓	✓
Run-time diagnostics	✓	✓	✓	✓	✓	Most	Most	Most
Storage Protection Keys	-	✓	✓	✓	✓	-	-	-
Dynamic Trace	✓	✓	✓	✓	✓	-	-	✓
Operating System FFDC	-	✓	✓	✓	✓	-	-	-
Error log analysis	✓	✓	✓	✓	✓	✓	✓	✓
Service Processor support for:								
Built-in-Self-Tests (BIST) for logic and arrays	✓	✓	✓	✓	✓	✓	✓	✓
Wire tests	✓	✓	✓	✓	✓	✓	✓	✓

RAS feature	AIX 5.3	AIX 6.1	AIX 7.1	IBM i 6.1.1	IBM i 7.1	RHEL 5	SLES 10 SP 3	SLES 11 SP 1
Component initialization	✓	✓	✓	✓	✓	✓	✓	✓
Serviceability								
Boot-time progress indicators	✓	✓	✓	✓	✓	Most	Most	Most
Firmware error codes	✓	✓	✓	✓	✓	✓	✓	✓
Operating system error codes	✓	✓	✓	✓	✓	Most	Most	Most
Inventory collection	✓	✓	✓	✓	✓	✓	✓	✓
Environmental and power warnings	✓	✓	✓	✓	✓	✓	✓	✓
Hot-plug fans, power supplies	✓	✓	✓	✓	✓	✓	✓	✓
Extended error data collection	✓	✓	✓	✓	✓	✓	✓	✓
SP “call home” on non-HMC configurations	✓	✓	✓	✓	✓	✓	✓	✓
I/O drawer redundant connections	✓	✓	✓	✓	✓	✓	✓	✓
I/O drawer hot add and concurrent repair	✓	✓	✓	✓	✓	✓	✓	✓
Concurrent RIO/GX adapter add	✓	✓	✓	✓	✓	✓	✓	✓
Concurrent cold-repair of GX adapter	✓	✓	✓	✓	✓	✓	✓	✓
Concurrent add of powered I/O rack to Power 795	✓	✓	✓	✓	✓	✓	✓	✓
SP mutual surveillance w/ POWER Hypervisor	✓	✓	✓	✓	✓	✓	✓	✓
Dynamic firmware update with HMC	✓	✓	✓	✓	✓	✓	✓	✓
Service Agent Call Home Application	✓	✓	✓	✓	✓	✓	✓	✓
Guiding light LEDs	✓	✓	✓	✓	✓	✓	✓	✓
Lightpath LEDs	✓	✓	✓	✓	✓	✓	✓	✓
System dump for memory, POWER Hypervisor, SP	✓	✓	✓	✓	✓	✓	✓	✓
Infocenter / Systems Support Site service publications	✓	✓	✓	✓	✓	✓	✓	✓
System Support Site education	✓	✓	✓	✓	✓	✓	✓	✓
Operating system error reporting to HMC SFP	✓	✓	✓	✓	✓	✓	✓	✓
RMC secure error transmission subsystem	✓	✓	✓	✓	✓	✓	✓	✓
Health check scheduled operations with HMC	✓	✓	✓	✓	✓	✓	✓	✓
Operator panel (real or virtual)	✓	✓	✓	✓	✓	✓	✓	✓
Concurrent operator panel maintenance	✓	✓	✓	✓	✓	✓	✓	✓
Redundant HMCs	✓	✓	✓	✓	✓	✓	✓	✓
Automated server recovery/restart	✓	✓	✓	✓	✓	✓	✓	✓
High availability clustering support	✓	✓	✓	✓	✓	✓	✓	✓
Repair and Verify Guided Maintenance	✓	✓	✓	✓	✓	Most	Most	Most
Concurrent kernel update	-	✓	✓	✓	✓	-	-	-

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

For information about ordering these publications, see “How to get Redbooks publications” on page 143. Note that various documents referenced here might be available in softcopy only.

- *IBM Power 720 and 740 Technical Overview and Introduction*, REDP-4637
- *IBM Power 795 Technical Overview and Introduction*, REDP-4640
- *PowerVM Virtualization on IBM System p: Introduction and Configuration Fourth Edition*, SG24-7940
- *IBM PowerVM Virtualization Managing and Monitoring*, SG24-7590
- *IBM PowerVM Live Partition Mobility*, SG24-7460
- *IBM System p Advanced POWER Virtualization (PowerVM) Best Practices*, REDP-4194
- *PowerVM Migration from Physical to Virtual Storage*, SG24-7825
- *IBM System Storage DS8000: Copy Services in Open Environments*, SG24-6788
- *IBM System Storage DS8700 Architecture and Implementation*, SG24-8786
- *PowerVM and SAN Copy Services*, REDP-4610
- *SAN Volume Controller V4.3.0 Advanced Copy Services*, SG24-7574

Other publications

These publications are also relevant as further information sources:

- IBM Power Systems Facts and Features POWER7 Blades and Servers:
  ftp://public.dhe.ibm.com/common/ssi/ecm/en/pob03022usen/POB03022USEN.PDF
- 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:
  ftp://public.dhe.ibm.com/common/ssi/ecm/en/pod03048usen/POD03048USEN.PDF
- IBM Power 720 server Data Sheet:
  ftp://public.dhe.ibm.com/common/ssi/ecm/en/pod03049usen/POD03049USEN.PDF
- IBM Power 730 server Data Sheet:
  ftp://public.dhe.ibm.com/common/ssi/ecm/en/pod03050usen/POD03050USEN.PDF
- IBM Power 740 server Data Sheet:
  ftp://public.dhe.ibm.com/common/ssi/ecm/en/pod03051usen/POD03051USEN.PDF

- IBM Power 750 server Data Sheet:

  http://www.ibm.com/common/ssi/fcgi-bin/ssialias?infotype=PM&subtype=SP&appname=STGE_PO_PO_USEN&htmlfid=POD03034USEN&attachment=POD03034USEN.PDF

- IBM Power 755 server Data Sheet:

  http://www.ibm.com/common/ssi/fcgi-bin/ssialias?infotype=PM&subtype=SP&appname=STGE_PO_PO_USEN&htmlfid=POD03035USEN&attachment=POD03035USEN.PDF

- IBM Power 770 server Data Sheet:

  http://www.ibm.com/common/ssi/fcgi-bin/ssialias?infotype=PM&subtype=SP&appname=STGE_PO_PO_USEN&htmlfid=POD03031USEN&attachment=POD03031USEN.PDF

- IBM Power 780 server Data Sheet:

  http://www.ibm.com/common/ssi/fcgi-bin/ssialias?infotype=PM&subtype=SP&appname=STGE_PO_PO_USEN&htmlfid=POD03032USEN&attachment=POD03032USEN.PDF

- IBM Power 795 server Data Sheet:

  ftp://public.dhe.ibm.com/common/ssi/ecm/en/pod03053usen/POD03053USEN.PDF

- Active Memory Expansion: Overview and Usage Guide:

  http://www-03.ibm.com/systems/power/hardware/whitepapers/am_exp.html

- Migration combinations of processor compatibility modes for active Partition Mobility:

  http://publib.boulder.ibm.com/infocenter/powersys/v3r1m5/topic/p7hc3/iphc3pcmcombosact.htm

- Advance Toolchain for Linux website:

  http://www.ibm.com/developerworks/wikis/display/hpccentral/How+to+use+Advance+Toolchain+for+Linux+on+POWER

Online resources

These websites are also relevant as further information sources:

- IBM Power Systems Hardware Information Center:

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

- IBM System Planning Tool website:

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

- Download from the 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/systems/support

- IBM Electronic Services portal:

  https://www.ibm.com/support/electronic/portal

- IBM Storage website:

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

How to get Redbooks publications

You can search for, view, or download Redbooks publications, Redpapers publications, Technotes, draft publications and Additional materials, as well as order hardcopy Redbooks publications, at this website:

ibm.com/redbooks

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

Redpaper™

Features the POWER7 processor providing advanced multi-core technology

PowerVM, enterprise-level RAS all in an entry server package

2U rack-mount design for midrange performance

This IBM Redpaper publication is a comprehensive guide covering the IBM Power 710 and Power 730 servers supporting AIX, IBM i, and Linux operating systems. The goal of this paper is to introduce the major innovative Power 710 and 730 offerings and their prominent functions, including these:

- The POWER7 processor available at frequencies of 3.0 GHz, 3.55 GHz, and 3.7 GHz
- The specialized POWER7 Level 3 cache that provides greater bandwidth, capacity, and reliability
- The 1 Gb or 10 Gb Integrated Virtual Ethernet adapter, included with each server configuration, and providing native hardware virtualization
- PowerVM virtualization including PowerVM Live Partition Mobility and PowerVM Active Memory Sharing
- Active Memory Expansion that provides more usable memory than what is physically installed on the system
- EnergyScale technology that provides features such as power trending, power-saving, capping of power, and thermal measurement

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-4636-00

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