

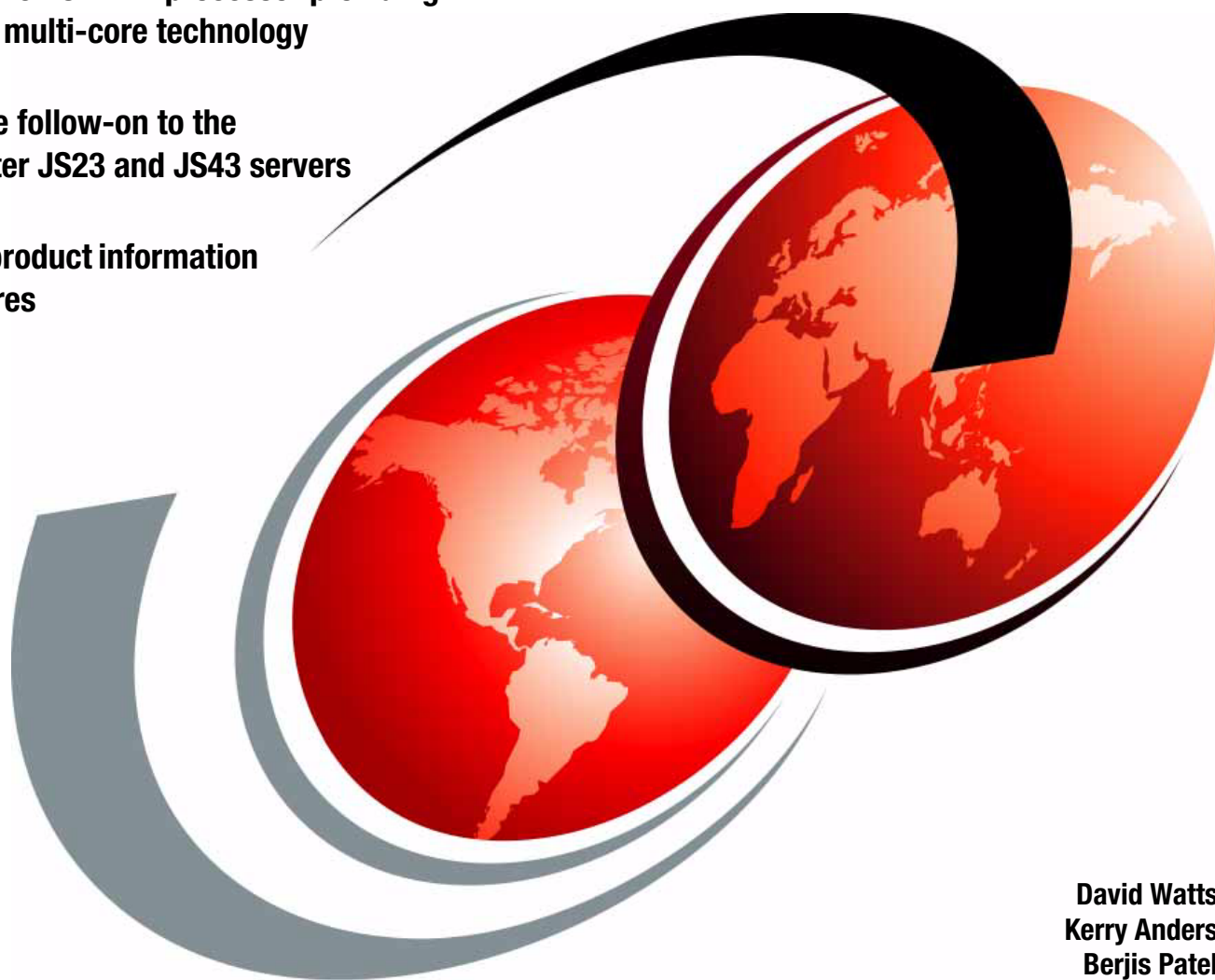
IBM BladeCenter PS700, PS701, and PS702

Technical Overview and Introduction

Features the POWER7 processor providing advanced multi-core technology

Details the follow-on to the BladeCenter JS23 and JS43 servers

Includes product information and features



David Watts
Kerry Anders
Berjis Patel



International Technical Support Organization

**IBM BladeCenter PS700, PS701, and PS702 Technical
Overview and Introduction**

May 2010

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

First Edition (May 2010)

This edition applies to:

- ▶ IBM BladeCenter PS700, 8406-70Y
- ▶ IBM BladeCenter PS701, 8406-71Y
- ▶ IBM BladeCenter PS702, 8406-71Y + FC 8358

This document created or updated on August 30, 2011.

© Copyright International Business Machines Corporation 2010. All rights reserved.

Note to U.S. Government Users Restricted Rights -- Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.

Contents

| | |
|---|------|
| Notices | vii |
| Trademarks | viii |
| Preface | ix |
| The team who wrote this paper | ix |
| Now you can become a published author, too! | x |
| Comments welcome | xi |
| Stay connected to IBM Redbooks | xi |
| Chapter 1. Introduction and general description | 1 |
| 1.1 Overview of PS700, PS701, and PS702 blade servers | 2 |
| 1.2 IBM BladeCenter support | 4 |
| 1.2.1 Supported BladeCenter chassis | 4 |
| 1.2.2 Number of PS700, PS701, and PS702 blades in a chassis | 14 |
| 1.3 Operating environment | 14 |
| 1.4 Physical package | 15 |
| 1.5 System features | 16 |
| 1.5.1 PS700 system features | 16 |
| 1.5.2 PS701 system features | 18 |
| 1.5.3 PS702 system features | 20 |
| 1.5.4 Minimum features for the POWER7 processor-based blade servers | 21 |
| 1.5.5 Power supply features | 22 |
| 1.5.6 Processor | 22 |
| 1.5.7 Memory features | 23 |
| 1.5.8 I/O features | 24 |
| 1.5.9 Disk features | 28 |
| 1.5.10 Standard onboard features | 28 |
| 1.6 Supported BladeCenter I/O modules | 29 |
| 1.6.1 Ethernet switch and intelligent pass through modules | 30 |
| 1.6.2 SAS I/O modules | 31 |
| 1.6.3 Fibre Channel switch and pass-thru modules | 31 |
| 1.6.4 Converged networking I/O modules | 32 |
| 1.6.5 InfiniBand switch module | 33 |
| 1.6.6 Multi-switch Interconnect Module | 33 |
| 1.6.7 Multi-switch Interconnect Module for BladeCenter HT | 34 |
| 1.7 Comparison between PS700, PS701, PS702, and 750 models | 35 |
| 1.8 Building to order | 36 |
| 1.9 Model upgrades | 36 |
| Chapter 2. Architecture and technical overview | 37 |
| 2.1 Architecture | 38 |
| 2.2 The IBM POWER7 processor | 38 |
| 2.2.1 POWER7 processor overview | 39 |
| 2.2.2 POWER7 processor core | 40 |
| 2.2.3 Simultaneous multithreading | 41 |
| 2.2.4 Memory access | 42 |
| 2.2.5 Flexible POWER7 processor packaging and offerings | 42 |
| 2.2.6 On-chip L3 cache innovation and intelligent cache | 43 |
| 2.2.7 POWER7 processor and intelligent energy | 45 |

| | |
|--|-----------|
| 2.2.8 Comparison of the POWER7 and POWER6 processors | 45 |
| 2.3 POWER7 processor-based blades | 46 |
| 2.4 Memory subsystem | 46 |
| 2.4.1 Memory placement rules | 46 |
| 2.5 Technical comparison | 51 |
| 2.6 Internal I/O subsystem | 52 |
| 2.6.1 Peripheral Component Interconnect Express (PCIe) bus | 52 |
| 2.6.2 PCIe slots | 52 |
| 2.6.3 I/O expansion cards | 58 |
| 2.6.4 Embedded SAS Controller | 60 |
| 2.6.5 HEA ports | 60 |
| 2.6.6 Embedded USB controller | 60 |
| 2.7 Integrated Virtual Ethernet | 61 |
| 2.7.1 IVE subsystem | 62 |
| 2.8 Service processor | 63 |
| 2.8.1 Server console access by SOL | 63 |
| 2.9 Internal storage | 65 |
| 2.9.1 Hardware RAID function | 68 |
| 2.9.2 External SAS connections | 68 |
| 2.10 External disk subsystems | 68 |
| 2.10.1 IBM BladeCenter S Disk Storage Modules | 69 |
| 2.10.2 IBM System Storage | 69 |
| 2.11 IVM | 71 |
| 2.12 Operating system support | 72 |
| 2.13 IBM EnergyScale | 74 |
| 2.13.1 IBM EnergyScale technology | 74 |
| 2.13.2 EnergyScale device | 76 |
| Chapter 3. Virtualization | 77 |
| 3.1 POWER Hypervisor | 78 |
| 3.2 POWER processor modes | 82 |
| 3.3 PowerVM | 83 |
| 3.3.1 PowerVM Editions | 83 |
| 3.3.2 Logical partitions | 85 |
| 3.3.3 VIOS | 88 |
| 3.3.4 PowerVM Lx86 | 92 |
| 3.3.5 PowerVM Live Partition Mobility | 93 |
| 3.3.6 Active Memory Sharing | 95 |
| 3.3.7 N_Port ID Virtualization (NPIV) | 96 |
| 3.3.8 Supported PowerVM features by operating system | 98 |
| Chapter 4. Continuous availability and manageability | 99 |
| 4.1 Introduction | 100 |
| 4.2 Reliability | 100 |
| 4.2.1 Designed for reliability | 101 |
| 4.2.2 Placement of components | 101 |
| 4.2.3 Redundant components and concurrent repair | 101 |
| 4.3 Availability | 101 |
| 4.3.1 Partition availability priority | 102 |
| 4.3.2 General detection and deallocation of failing components | 103 |
| 4.3.3 Memory protection | 104 |
| 4.3.4 Cache protection | 107 |
| 4.3.5 Special uncorrectable error handling | 108 |

| | |
|--|------------|
| 4.3.6 PCI extended error handling | 108 |
| 4.4 Serviceability | 109 |
| 4.4.1 Detecting | 110 |
| 4.4.2 Diagnosing | 112 |
| 4.4.3 Reporting | 113 |
| 4.4.4 Notifying the client. | 116 |
| 4.4.5 Locating and servicing parts requiring service | 117 |
| 4.5 Manageability | 120 |
| 4.5.1 Service user interfaces | 120 |
| 4.5.2 IBM Power Systems firmware maintenance | 122 |
| 4.5.3 Electronic Service Agent tool | 124 |
| 4.5.4 BladeCenter Service Advisor | 125 |
| Abbreviations and acronyms | 129 |
| Related publications | 131 |
| IBM Redbooks | 131 |
| Other publications | 131 |
| Online resources | 132 |
| How to get Redbooks | 132 |
| Help from IBM | 132 |

Notices

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

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

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785 U.S.A.

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

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

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

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

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](http://www.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:

| | | |
|---|-------------------------|---|
| AIX 5L™ | IBM® | pSeries® |
| AIX® | Micro-Partitioning™ | Redbooks® |
| BladeCenter® | POWER Hypervisor™ | Redpaper™ |
| Calibrated Vecteded Cooling™ | Power Systems™ | Redbooks (logo)  ® |
| DS4000® | Power Systems Software™ | ServerProven® |
| DS8000® | POWER4™ | System p5® |
| Electronic Service Agent™ | POWER5™ | System p® |
| EnergyScale™ | POWER6+™ | System Storage® |
| FlashCopy® | POWER6® | System x® |
| Focal Point™ | POWER7™ | System z® |
| IBM Systems Director Active Energy Manager™ | PowerVM™ | Tivoli® |
| | POWER® | XIV® |

The following terms are trademarks of other companies:

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

Microsoft, Windows, 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

The IBM® BladeCenter® PS700, PS701, and PS702 are premier blades for 64-bit applications. They are designed to minimize complexity, improve efficiency, automate processes, reduce energy consumption, and scale easily. These blade servers are based on the IBM POWER7™ processor and support AIX®, IBM i, and Linux® operating systems. Their ability to coexist in the same chassis with other IBM BladeCenter blades servers enhances the ability to deliver rapid return of investment demanded by clients and businesses.

This IBM Redpaper™ is a comprehensive guide covering the IBM BladeCenter PS700, PS701, and PS702 servers. The goal of this paper is to introduce the offerings and their prominent features and functions.

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, Poughkeepsie Center.

David Watts is a Consulting IT Specialist at the IBM ITSO Center in Raleigh. He manages residencies and produces IBM Redbooks® publications on hardware and software topics related to IBM BladeCenter and IBM System x® servers and associated client platforms. He has authored over 80 books, papers, and Web docs. He has worked for IBM both in the US and Australia since 1989. David is an IBM Certified IT Specialist and a member of the IT Specialist Certification Review Board. He holds a Bachelor of Engineering degree from the University of Queensland (Australia)

Kerry Anders is a Consultant in System p® Lab Services for the IBM Systems and Technology Group, based in Austin, Texas. He supports clients in implementing IBM Power Systems™ blades using Virtual I/O Server, Integrated Virtualization Manager, and AIX. Prior IBM Redbooks publication projects include the *IBM BladeCenter JS12 and JS22 Implementation Guide*, SG24-7655 and the *IBM BladeCenter JS23 and JS43 Implementation Guide*, SG24-7740. Previously, he was the Systems Integration Test Team Lead for the IBM BladeCenter JS21blade with IBM SAN storage using AIX and Linux. His prior work includes test experience with the JS20 blade, also using AIX and Linux in SAN environments. Kerry began his career with IBM in the Federal Systems Division supporting NASA at the Johnson Space Center as a Systems Engineer. He transferred to Austin in 1993.

Berjis Patel is a Senior IT Specialist with System Sales Implementation Services with IBM Global Technology Services in Canada. He has over 20 years of experience in the IT industry with more than 15 years with IBM UNIX® (AIX) solutions. He is a certified presales specialist for IBM System p and has multiple IBM Hundred Percent Club awards. His area of expertise is consulting, selling, and implementing services such as consolidation, virtualization, migration, high-availability, and systems management solutions on IBM Power Systems. He has worked in various IBM locations including India, the Middle East, the USA, and now in Canada with different roles since 1995.



The team (l-r): David, Berjis, and Kerry

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

From IBM Power Systems development:

- ▶ Chris Austen
- ▶ Larry Cook
- ▶ Jeff Franke
- ▶ Tom Flynn
- ▶ Kaena Freitas
- ▶ Ghadir Gholami
- ▶ Jim Jordan
- ▶ Richard Lary
- ▶ Gregory McIntire
- ▶ Todd Rosedahl
- ▶ Steven Royer
- ▶ Mark Smolen
- ▶ Chris Sturgill
- ▶ Mike Stys

From IBM Power Systems marketing:

- ▶ John Biebelhausen
- ▶ Guy Paradise

From IBM Systems & Technology Group:

- ▶ Michael L. Nelson

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 e-mail to:

redbooks@us.ibm.com

- ▶ Mail your comments to:

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

Stay connected to IBM Redbooks

- ▶ Find us on Facebook:

<http://www.facebook.com/IBMRedbooks>

- ▶ Follow us on twitter:

<http://twitter.com/ibmredbooks>

- ▶ Look for us on LinkedIn:

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

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

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

- ▶ Stay current on recent Redbooks publications with RSS Feeds:

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



Introduction and general description

This chapter provides an introduction and general description to the new IBM BladeCenter POWER7 processor-based blade servers. These new blades offer processor scalability from four cores to 16 cores:

- ▶ IBM BladeCenter PS700: Single-wide blade with a single-socket 4-core processor
- ▶ IBM BladeCenter PS701: Single-wide blade with a single-socket 8-core processor
- ▶ IBM BladeCenter PS702: Double-wide blade with two single-socket 8-core processors

The new PS700, PS701, and PS702 blades are premier blades for 64-bit applications. They are designed to minimize complexity, improve efficiency, automate processes, reduce energy consumption, and scale easily.

The POWER7 processor-based PS700, PS701, and PS702 blades support AIX, IBM i, and Linux operating systems. Their ability to coexist in the same chassis with other IBM BladeCenter blades servers enhances the ability to deliver rapid return of investment demanded by clients and businesses.

This chapter covers the following topics:

- ▶ 1.1, “Overview of PS700, PS701, and PS702 blade servers” on page 2
- ▶ 1.2, “IBM BladeCenter support” on page 4
- ▶ 1.3, “Operating environment” on page 14
- ▶ 1.4, “Physical package” on page 15
- ▶ 1.5, “System features” on page 16
- ▶ 1.6, “Supported BladeCenter I/O modules” on page 29
- ▶ 1.7, “Comparison between PS700, PS701, PS702, and 750 models” on page 35
- ▶ 1.8, “Building to order” on page 36
- ▶ 1.9, “Model upgrades” on page 36

1.1 Overview of PS700, PS701, and PS702 blade servers

Figure 1-1 shows the IBM BladeCenter PS700, PS701, and PS702 blade servers.



Figure 1-1 The IBM BladeCenter PS702, BladeCenter PS701, and BladeCenter PS700

The PS700 blade server

The PS700 Blade Server (8406-70Y) is a single socket, single wide 4-core 3.0 GHz POWER7 processor-based server. The POWER7 processor is a 64-bit, 4-core with 256 KB L2 cache per core and 4 MB L3 cache per core.

The PS700 blade server has eight DDR3 memory DIMM slots. The industry standard VLP DDR3 Memory DIMMs are either 4 GB or 8 GB running at 1066 MHz. The memory is supported in pairs, thus the minimum memory required for PS700 blade server is 8 GB (two 4 GB DIMMs). The maximum memory that can be supported is 64 GB (eight 8 GB DIMMs).

It has two Host Ethernet Adapters (HEA) 1 GB integrated Ethernet ports that are connected to the BladeCenter chassis fabric (midplane). The PS700 has an integrated SAS controller that supports local (on-board) storage, integrated USB controller and Serial over LAN console access through the service processor, and the BladeCenter Advance Management Module.

It supports two on-board disk drive bays. The on-board storage can be one or two 2.5-inch SAS HDD. The integrated SAS controller supports RAID 0, RAID 1, and RAID 10 hardware when two HDDs are used.

The PS700 also supports one PCIe CIOv expansion card slot and one PCIe CFFh expansion card slot. See 1.5.8, “I/O features” on page 24 for supported I/O expansion cards.

The PS701 blade server

The PS701 blade server (8406-71Y) is a single socket, single-wide 8-core 3.0 GHz POWER7 processor-based server. The POWER7 processor is a 64-bit, 8-core with 256 KB L2 cache per core and 4 MB L3 cache per core.

The PS701 blade server has 16 DDR3 memory DIMM slots. The industry standard VLP DDR3 memory DIMMs are either 4 GB or 8 GB running at 1066 MHz. The memory is supported in pairs, thus the minimum memory required for PS701 blade server is 8 GB (two 4 GB DIMMs). The maximum memory that can be supported is 128 GB (16x 8 GB DIMMs).

The PS701 blade server has two Host Ethernet Adapters (HEA) 1 GB integrated Ethernet ports that are connected to the BladeCenter chassis fabric (midplane). The PS701 also has an integrated SAS controller that supports local (on-board) storage, integrated USB controller and Serial over LAN console access through the service processor, and the BladeCenter Advance Management Module.

The PS701 has one on-board disk drive bay. The on-board storage can be one 2.5-inch SAS HDD. The PS701 also supports one PCIe CIOv expansion card slot and one PCIe CFFh expansion card slot. See 1.5.8, “I/O features” on page 24 for supported I/O expansion cards.

The PS702 blade server

The PS702 blade server (8406-71Y +FC 8358) is a two socket, double-wide 16-core 3.0 GHz POWER7 processor-based server. The POWER7 processor is a 64-bit, 8-core with 256 KB L2 cache per core and 4 MB L3 cache per core. The PS702 combines a single-wide base blade (PS701) and an expansion unit (feature 8358), referred to as double-wide blade, which occupies two adjacent slots in the IBM BladeCenter chassis.

The PS702 blade server has 32 DDR3 memory DIMM slots. The industry standard VLP DDR3 memory DIMMs are either 4 GB or 8 GB running at 1066 MHz. The memory is supported in pairs, thus the minimum memory required for PS702 blade server is 8 GB (two 4 GB DIMMs). The maximum memory that can be supported is 256 GB (32x 8 GB DIMMs).

Note: The PS702 blade server can have a minimum of 8 GB memory based as per architecture, but we recommend a reasonable ratio between cores and memory.

The PS702 blade server has four Host Ethernet Adapter 1 GB integrated Ethernet ports that are connected to the BladeCenter chassis fabric (midplane). The PS702 also has an integrated SAS controller that supports local (on-board) storage, integrated USB controller and Serial over LAN console access through the service processor, and the BladeCenter Advance Management Module.

The PS702 blade server has two disk drive bays, one on the base blade and one on the expansion unit. The on-board storage can be one or two 2.5-inch SAS disk drives. The integrated SAS controller supports RAID 0, RAID 1 and RAID 10 hardware when two HDDs are used.

The PS702 supports two PCIe CIOv expansion card slot and two PCIe CFFh expansion card slots. See 1.5.8, “I/O features” on page 24 for supported I/O expansion cards.

Note: For the PS702 blade server, the service processor (FSP or just SP) in the expansion blade is set to IO mode, which provides control busses from IOs, but does not provide redundancy and backup operational support to the SP in the base blade.

1.2 IBM BladeCenter support

Blade servers are thin servers that insert into a single rack-mounted chassis that supplies shared power, cooling, and networking infrastructure. Each server is an independent server with its own processors, memory, storage, network controllers, operating system, and applications. The IBM BladeCenter chassis is the container for the blade servers and shared infrastructure devices.

The IBM BladeCenter chassis can contain a mix of POWER®, Intel®, Cell and AMD processor-based blades. Depending on the IBM BladeCenter chassis selected, combinations of Ethernet, SAS, Fibre Channel, and FCoE I/O fabrics can also be shared within the same chassis.

All chassis can offer full redundancy for all shared infrastructure, network, and I/O fabrics. Having multiple power supplies, network switches, and I/O switches contained within a BladeCenter chassis eliminates single points of failure in these areas.

The following sections describe the BladeCenter chassis that support the PS700, PS701, and PS702 blades. For a comprehensive look at all aspects of BladeCenter products see the IBM Redbooks publication, *IBM BladeCenter Products and Technology*, SG24-7523, available from the following Web page:

<http://www.redbooks.ibm.com/abstracts/sg247523.html>

1.2.1 Supported BladeCenter chassis

The PS700, PS701, and PS702 blades are supported in the IBM BladeCenter chassis as listed in Table 1-1.

Table 1-1 The blade servers supported in each BladeCenter chassis

| Blade | Machine type-model | Blade width | BC S 8886 | BC E 8677 ^a | BC T 8720 | BC T 8730 | BC H 8852 | BC HT 8740 | BC HT 8750 |
|-------|--------------------|-------------|-----------|------------------------|-----------|-----------|-----------|------------|------------|
| PS700 | 8406-70Y | 1 slot | Yes | Yes ^b | No | No | Yes | Yes | Yes |
| PS701 | 8406-71Y | 1 slot | Yes | No | No | No | Yes | Yes | Yes |
| PS702 | 8406-71Y | 2 slot | Yes | No | No | No | Yes | Yes | Yes |

a. BladeCenter E requires an Advanced Management Module and a minimum of two 2000 watt power supplies.

b. Only specific models of the BladeCenter E support the PS700. See Table 1-2.

A detailed description of each supported BladeCenter for the PS700, PS701, and PS702 blades is contained in the following sections.

- ▶ IBM BladeCenter E (PS700 only support) provides the greatest density and common fabric support. It is the lowest entry cost option. See “BladeCenter E” on page 5 for details on the chassis.

Only specific models of the BladeCenter E models listed are supported as shown in Table 1-2.

Table 1-2 BladeCenter E models that support the PS700

| BladeCenter E models ^a | Supports the PS700 |
|-----------------------------------|--------------------|
| 8677-3Xx | No |
| 8677-3Rx | No |
| 8677-E2x | No |
| 8677-3Sx | Yes ^b |
| 8677-4Sx | Yes |
| 8677-3Tx | Yes ^b |
| 8677-4Tx | Yes |

a. x = country-specific letter (for example, EMEA MTM is 8677-4SG, and the US MTM is 8677-4SU).

b. The 3Sx and 3Tx models are supported but only with upgraded (2320W) power supplies

- ▶ IBM BladeCenter H delivers high performance, extreme reliability, and ultimate flexibility for the most demanding IT environments. See “BladeCenter H” on page 7.
- ▶ IBM BladeCenter HT models are designed for high-performance flexible telecommunications environments by supporting high-speed internet working technologies (such as 10G Ethernet). They provide a robust platform for NGNs. See “BladeCenter HT” on page 9.
- ▶ IBM BladeCenter S combines the power of blade servers with integrated storage, all in an easy-to-use package designed specifically for the office and distributed enterprise environments. See “BladeCenter S” on page 12.

Note: The number of blade servers that can be installed into chassis is dependent on the power supply configuration, power supply input (110V/208V BladeCenter S only) and power domain configuration options. See 1.2.2, “Number of PS700, PS701, and PS702 blades in a chassis” on page 14 for more information.

BladeCenter E

IBM designed the IBM BladeCenter E (machine type 8677) to be a highly modular chassis to accommodate a range of diverse business requirements. BladeCenter supports not only blade servers, but also a wide range of networking modules, including Gigabit Ethernet, Fibre Channel, and SAS for connectivity to the client’s existing network environment. BladeCenter E also supports a redundant pair of Management Modules for comprehensive systems management.

Providing a wide selection of integrated switching options, BladeCenter systems lower the total cost of ownership (TCO) by eliminating the need to purchase additional keyboards, videos, and mice (KVM), Ethernet and Fibre Channel switches, or the cumbersome and expensive cabling required by the switches. BladeCenter is a leader in the industry in providing flexibility and a variety of integration choices with components that fit your infrastructure and deliver a comprehensive blade solution.

BladeCenter E’s superior density and feature set are made possible by the BladeCenter E innovative chassis architecture. Because BladeCenter E uses super energy-efficient components and shared infrastructure architecture, clients realize lower power consumption when compared to their most likely alternative (that is, non-blade server designs).

BladeCenter E's lower power consumption and Calibrated Vecteded Cooling™ allow more servers to fit in a tight power or cooling environment.

Figure 1-2 displays the front view of an IBM BladeCenter E.



Figure 1-2 BladeCenter E front view

Figure 1-3 displays the rear view of an IBM BladeCenter E.



Figure 1-3 BladeCenter E rear view

The key features of IBM BladeCenter E chassis are as follows:

- ▶ A rack-optimized, 7 U modular design enclosure for up to 14 hot-swap blades
- ▶ A high-availability mid-plane that supports hot-swap of individual blades
- ▶ For 8677-3Sx, two 2,000-watt, hot-swap power modules and support for two optional 2,000-watt power modules, offering redundancy and power for robust configurations

- ▶ For 8677-4Sx, two 2,320-watt, hot-swap power modules and support for two optional 2,320-watt power modules, offering higher power and performance than previous models, extending support to a wider range of blades
- ▶ Two hot-swap blowers
- ▶ An Advanced Management Module that provides chassis-level solutions, simplifying deployment and management of your installation
- ▶ Support for up to four network or storage switches or pass-through modules
- ▶ A light path diagnostic panel, and USB 2.0 port
- ▶ Support for UltraSlim enhanced SATA DVD-ROM and multi-burner drives
- ▶ IBM Systems Director and Tivoli® Provisioning Manager for OS Deployments for easy installation and management
- ▶ Energy-efficient design and innovative features to maximize productivity and reduce power usage
- ▶ Extreme density and integration to ease data center space constraints
- ▶ Help in protecting your IT investment through IBM BladeCenter family longevity, compatibility, and innovation leadership in blades
- ▶ Support for the latest generation of IBM BladeCenter blades, providing investment protection

BladeCenter H

IBM BladeCenter H delivers high performance, extreme reliability, and ultimate flexibility to even the most demanding IT environments. In 9 U of rack space, the BladeCenter H chassis can contain up to 14 blade servers, 10 switch modules, and four power supplies to provide the necessary I/O network switching, power, cooling, and control panel information to support the individual servers.

The chassis supports up to four traditional fabrics using networking switches, storage switches, or pass through devices. The chassis also supports up to four high-speed fabrics for support of protocols such as 4X InfiniBand or 10 Gigabit Ethernet. The built-in media tray includes light path diagnostics, two front USB inputs, and an optical drive.

Figure 1-4 displays the front view of an IBM BladeCenter H.



Figure 1-4 BladeCenter H front view

Figure 1-5 displays the rear view of an IBM BladeCenter H.



Figure 1-5 BladeCenter H rear view

The key features of IBM BladeCenter H chassis are as follows:

- ▶ A rack-optimized, 9 U modular design enclosure for up to 14 hot-swap blades

- ▶ A high-availability mid-plane that supports hot-swap of individual blades
- ▶ Two 2,900-watt, hot-swap power modules and support for two optional 2,900-watt power modules, offering redundancy and power for robust configurations
- ▶ Two hot-swap redundant blowers, and six or 12 supplemental fans with power supplies
- ▶ An Advanced Management Module that provides chassis-level solutions, simplifying deployment and management of your installation
- ▶ Support for up to four network or storage switches or pass-through modules
- ▶ Support for up to four bridge modules
- ▶ A light path diagnostic panel, and two USB 2.0 ports
- ▶ Serial port breakout connector
- ▶ Support for UltraSlim Enhanced SATA DVD-ROM and Multi-Burner Drives
- ▶ IBM Systems Director and Tivoli Provisioning Manager for OS Deployments for easy installation and management
- ▶ Energy-efficient design and innovative features to maximize productivity and reduce power usage
- ▶ Density and integration to ease data center space constraints
- ▶ Help in protecting your IT investment through IBM BladeCenter family longevity, compatibility, and innovation leadership in blades
- ▶ Support for the latest generation of IBM BladeCenter blades, helping provide investment protection

BladeCenter HT

The IBM BladeCenter HT is a 12-server blade chassis designed for high-density server installations, typically for telecommunications use. It offers high performance with the support of 10 G Ethernet installations. This 12 U high chassis with DC or AC power supplies provides a cost-effective, high-performance, high-availability solution for telecommunication networks and other rugged non-telecommunications environments. IBM BladeCenter HT chassis is positioned for expansion, capacity, redundancy, and carrier-grade NEBS level 3/ETSI compliance in DC models.

BladeCenter HT provides a solid foundation for next-generation networks (NGN) enabling service providers to become on demand providers. Coupled with technological expertise within the enterprise data center, IBM makes use of the industry know-how of key business partners to deliver added value within service provider networks.

Figure 1-6 shows the front view of the BladeCenter HT.



Figure 1-6 BladeCenter HT front view

Figure 1-7 shows the rear view of the BladeCenter HT.



Figure 1-7 BladeCenter HT rear view

BladeCenter HT delivers rich telecommunications features and functionality, including integrated servers, storage and networking, fault-tolerant features, optional hot swappable redundant DC or AC power supplies and cooling, and built-in system management resources. The result is a Network Equipment Building Systems (NEBS-3) and ETSI-compliant server platform optimized for next-generation networks.

The following BladeCenter HT applications are suited for these servers:

- ▶ Network management and security
 - Network management engine
 - Internet cache engine
 - RSA encryption
 - Gateways
 - Intrusion detection
- ▶ Network infrastructure
 - Softswitch
 - Unified messaging
 - Gateway/Gatekeeper/SS7 solutions
 - VOIP services and processing
 - Voice portals
 - IP translation database

The key features of the BladeCenter HT are as follows:

- ▶ Support for up to 12 blade servers, compatible with the other chassis in the BladeCenter family
- ▶ Four standard and four high-speed I/O module bays, compatible with the other chassis in the BladeCenter family
- ▶ A media tray at the front with light path diagnostics, two USB 2.0 ports, and optional compact flash memory module support
- ▶ Two hot-swap management-module bays (one management module standard)
- ▶ Four hot-swap power-module bays (two power modules standard)
- ▶ New serial port for direct serial connection to installed blades
- ▶ Compliance with the NEBS 3 and ETSI core network specifications

BladeCenter S

The BladeCenter S chassis can hold up to six blade servers, and up to 12 hot-swap 3.5-inch SAS or SATA disk drives in just 7 U of rack space. It can also include up to four C14 950-watt / 1450-watt power supplies. The BladeCenter S offers the necessary I/O network switching, power, cooling, and control panel information to support the individual servers.

The IBM BladeCenter S is one of five chassis in the BladeCenter family. The BladeCenter S provides an easy IT solution to the small and medium office and to the distributed enterprise. Figure 1-8 shows the front view of IBM BladeCenter S.

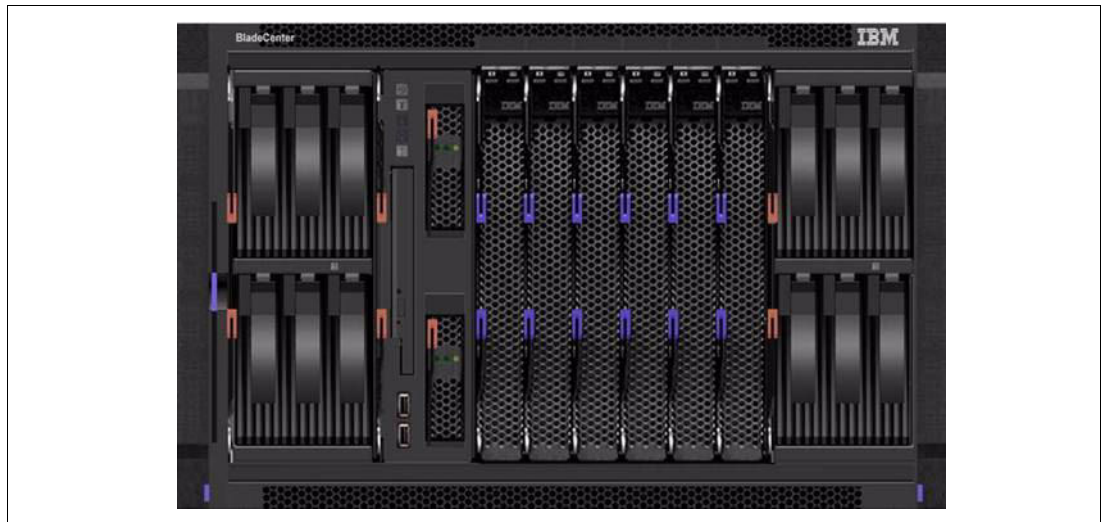


Figure 1-8 The front of the BladeCenter S chassis

Figure 1-9 shows the rear view of the chassis.



Figure 1-9 The rear of the BladeCenter S chassis

The key features of IBM BladeCenter S chassis are as follows:

- ▶ A rack-optimized, 7 U modular design enclosure for up to six hot-swap blades
- ▶ Two optional Disk Storage Modules for HDDs, six 3.5-inch SAS/SATA drives each
- ▶ High-availability mid-plane that supports hot-swap of individual blades
- ▶ Two 950/1450-watt, hot-swap power modules and support for two optional 950/1450-watt power modules, offering redundancy and power for robust configurations
- ▶ Four hot-swap redundant blowers, plus one fan in each power supply
- ▶ An Advanced Management Module that provides chassis-level solutions, simplifying deployment and management of your installation
- ▶ Support for up to four network or storage switches or pass-through modules
- ▶ A light path diagnostic panel, and two USB 2.0 ports
- ▶ Support for optional UltraSlim Enhanced SATA DVD-ROM and Multi-Burner Drives
- ▶ Support for SAS RAID Controller Module makes it easy for clients to buy the all-in-one BladeCenter S solution
- ▶ IBM Systems Director, Storage Configuration Manager (SCM), Start Now Advisor, and Tivoli Provisioning Manager for OS Deployments support for easy installation and management
- ▶ Energy-efficient design and innovative features to maximize productivity and reduce power usage
- ▶ Help in protecting your IT investment through IBM BladeCenter family longevity, compatibility, and innovation leadership in blades
- ▶ Support for the latest generation of IBM BladeCenter blades, helping provide investment protection

1.2.2 Number of PS700, PS701, and PS702 blades in a chassis

The number of POWER7 processor based blades that can be installed in a BladeCenter chassis depends on several factors:

- ▶ BladeCenter chassis type
- ▶ Number of power supplies installed
- ▶ Power supply voltage option (BladeCenter S only)
- ▶ BladeCenter power domain configuration

Table 1-3 shows the maximum number of PS700, PS701, and PS702 blades running in a maximum configuration (memory, disk, expansion cards) for each supported BladeCenter chassis that can be installed with fully redundant power and without performance reduction. IBM blades that are based on processor types other than POWER7 might reduce these numbers.

Table 1-3 PS700, PS701, and PS702 blades per chassis type

| | BladeCenter E ^a | | BladeCenter H | | BladeCenter HT | | BladeCenter S | | | |
|-------|----------------------------|------|----------------|------|----------------|------|---------------|------|--------|------|
| | 14 Slots Total | | 14 Slots Total | | 12 Slots Total | | 6 Slots Total | | | |
| | | | | | | | 110VAC | | 208VAC | |
| | 2 PS | 4 PS | 2 PS | 4 PS | 2 PS | 4 PS | 2 PS | 4 PS | 2 PS | 4 PS |
| PS700 | 6 | 14 | 7 | 14 | 6 | 12 | 2 | 6 | 2 | 6 |
| PS701 | None | None | 7 | 14 | 6 | 12 | 2 | 6 | 2 | 6 |
| PS702 | None | None | 3 | 7 | 3 | 6 | 1 | 3 | 1 | 3 |

a. BladeCenter E requires 2000 or 2300 watt power supplies

When mixing blades of different processor types in the same BladeCenter, the BladeCenter Power Configurator tool helps determine if the combination desired is valid. It is expected that this tool will be updated to include the PS700, PS701, and PS702 blade configurations. For more information about this update, see the following Web page:

<http://www.ibm.com/systems/bladecenter/powerconfig>

1.3 Operating environment

In this section, we list the operating environment specifications for the PS700, PS701, and PS702 blade servers and BladeCenter H and S.

IBM Blade Server PS700, PS701, and PS702

- ▶ Operating temperature
 - 10 to 35 °C (50 to 95 °F) at 0 to 914 meters altitude (0 to 3000 feet)
 - 10 to 32 °C (50 to 90 °F) at 914 to 2133 meters altitude (3000 to 7000 feet)
- ▶ Relative Humidity 8% to 80%
- ▶ Maximum Altitude 2133 meters (7000 ft.)

IBM BladeCenter H

- ▶ Operating temperature
 - 10.0 to 35 °C (50 to 95 °F) at 0 to 914 m (0 to 3000 ft.)
 - 10.0 to 32 °C (50 to 90 °F) at 914 to 2,133 m (3000 to 7000 ft.)
- ▶ Relative humidity 8% to 80%
- ▶ Maximum altitude: 2,133 meters (7000 ft.)

IBM BladeCenter S

- ▶ – Operating Temperature:
 - 10 to 35 °C (50° to 95°F) at 0 to 914 m (0 to 3000 ft.)
 - 10 to 32°C (50° to 90°F) at 914 to 2,133 m (3000 to 7000 ft.)
- ▶ Relative humidity: 8% to 80%
- ▶ Maximum altitude: 2,133 meters (7000 ft.)

IBM BladeCenter E

- ▶ Operating temperature
 - 10.0 to 35.0 °C (50 to 95 °F) at 0 to 914 m (0 to 3000 ft.)
 - 10.0 to 32.0 °C (50 to 90 °F) at 914 to 2133 m (3000 to 7000 ft.)
- ▶ Relative humidity: 8% to 80%
- ▶ Maximum altitude: 2133 meters (7000 ft.)

BladeCenter HT

- ▶ Operating temperature
 - 5 to 40 °C (41 to 104 °F) at -60 to 1800 m (-197 to 6000 ft.)
 - 5 to 30 °C (41 to 86 °F) at 1800m to 4000m (6000 to 13000 ft.)
- ▶ Relative humidity 5% to 85%
- ▶ Maximum altitude: 4000 meters (13000 ft.)

1.4 Physical package

The PS700, PS701 and the PS702 Blade Servers are supported in BladeCenter H, HT and S. BladeCenter E supports PS700 Blade Servers only.

This section describes the physical dimension of the POWER7 Blade Servers and the supported BladeCenter chassis only. Table 1-4 shows the physical dimensions of the PS700, PS701, and PS702 blade servers.

Table 1-4 Physical dimensions of PS700, PS701, and PS702 Blade Servers

| Dimension | PS700 blade server | PS701 blade server | PS702 blade server |
|-----------|--|--|--|
| Height | 9.65 inch (245 mm) | 9.65 inch (245 mm) | 9.65 inch (245 mm) |
| Width | 1.14 inch (29 mm) Single-wide blade | 1.14 inch (29 mm) Single-wide blade | 2.32 inch (59 mm) Double-wide blade |
| Depth | 17.55 inch (445 mm) | 17.55 inch (445 mm) | 17.55 inch (445 mm) |
| Weight | 9.6 lbs (4.35 kg) | 9.6 lbs (4.35 kg) | 19.2 lbs (8.7 kg) |

Table 1-5 shows the physical dimension of the BladeCenter chassis that supports the POWER7 processor based Blade Servers.

Table 1-5 Physical dimension of Supported BladeCenter chassis

| Dimension | BladeCenter H | BladeCenter S | BladeCenter E ^a | BladeCenter HT |
|-----------|-----------------|----------------|----------------------------|----------------|
| Height | 15.75" (400 mm) | 12" (305 mm) | 12" (305 mm) | 21" (528 mm) |
| Width | 17.4" (442 mm) | 17.5" (445 mm) | 17.5" (445 mm) | 17.4" (442 mm) |
| Depth | 28" (711 mm) | 28.9" (734 mm) | 28" (711 mm) | 27.8" (706 mm) |

a. PS700 only. The PS701 and PS702 are not supported in BladeCenter E chassis

1.5 System features

The PS700, PS701, and PS702 blade servers are 4-core, 8-core and 16-core POWER7 processor-based blade servers. This section describes the features on each of the POWER7 blade server. The following topics are covered:

- ▶ 1.5.1, "PS700 system features" on page 16
- ▶ 1.5.2, "PS701 system features" on page 18
- ▶ 1.5.3, "PS702 system features" on page 20
- ▶ 1.5.4, "Minimum features for the POWER7 processor-based blade servers" on page 21
- ▶ 1.5.5, "Power supply features" on page 22
- ▶ 1.5.6, "Processor" on page 22
- ▶ 1.5.7, "Memory features" on page 23
- ▶ 1.5.8, "I/O features" on page 24
- ▶ 1.5.9, "Disk features" on page 28
- ▶ 1.5.10, "Standard onboard features" on page 28

1.5.1 PS700 system features

The BladeCenter PS700, model 8406-70Y, is shown in Figure 1-10 on page 17. The features of the server are as follows:

- ▶ Machine type and model number
8406-70Y
- ▶ Form factor
Single-wide (30 mm) blade
- ▶ Processors:
 - Single-socket 4-core 64-bit POWER7 processor operating at a 3.0 GHz clock speed.
 - Based on CMOS 12S 45 nm SOI (silicon-on-insulator) technology,
 - Power consumption is 150w/socket
 - Single-wide (SW) Blade package
- ▶ Memory
 - 8 DIMM Slots
 - Minimum capacity 8 GB, maximum capacity 64 GB
 - Industry standard VLP DDR3 DIMMs

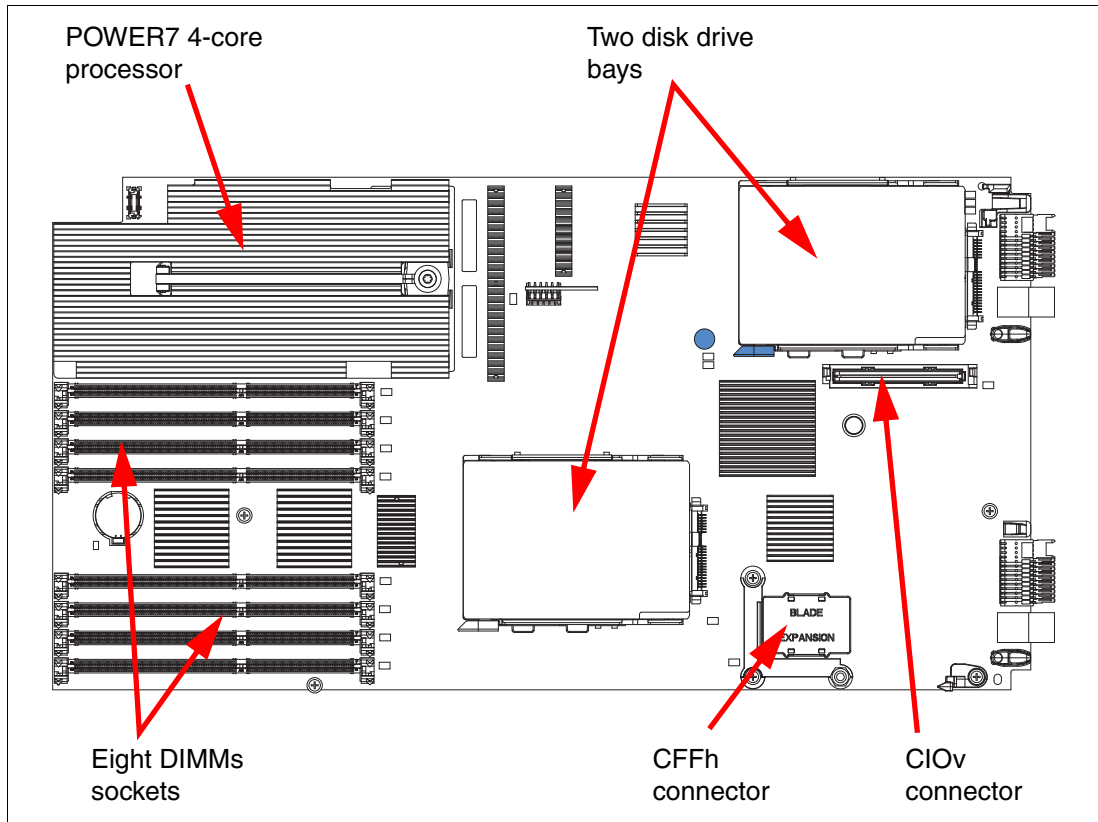


Figure 1-10 Top view of PS700 blade server

- ▶ Disk:
 - Two disk drive bays support one or two SAS HDD
 - Hardware mirroring RAID 0, RAID 1, and RAID 10
- ▶ On-board integrated features:
 - Service processor (SP)
 - Two 1 GB Ethernet ports (HEA)
 - SAS Controller
 - USB Controller that routes to the USB 2.0 port on the media tray
 - One Serial over LAN (SOL) Console through SP
- ▶ Expansion Card I/O Options:
 - One CIOv expansion card slot (PCIe)
 - One CFFh expansion card slot (PCIe)

1.5.2 PS701 system features

The BladeCenter PS701 is shown in Figure 1-11.

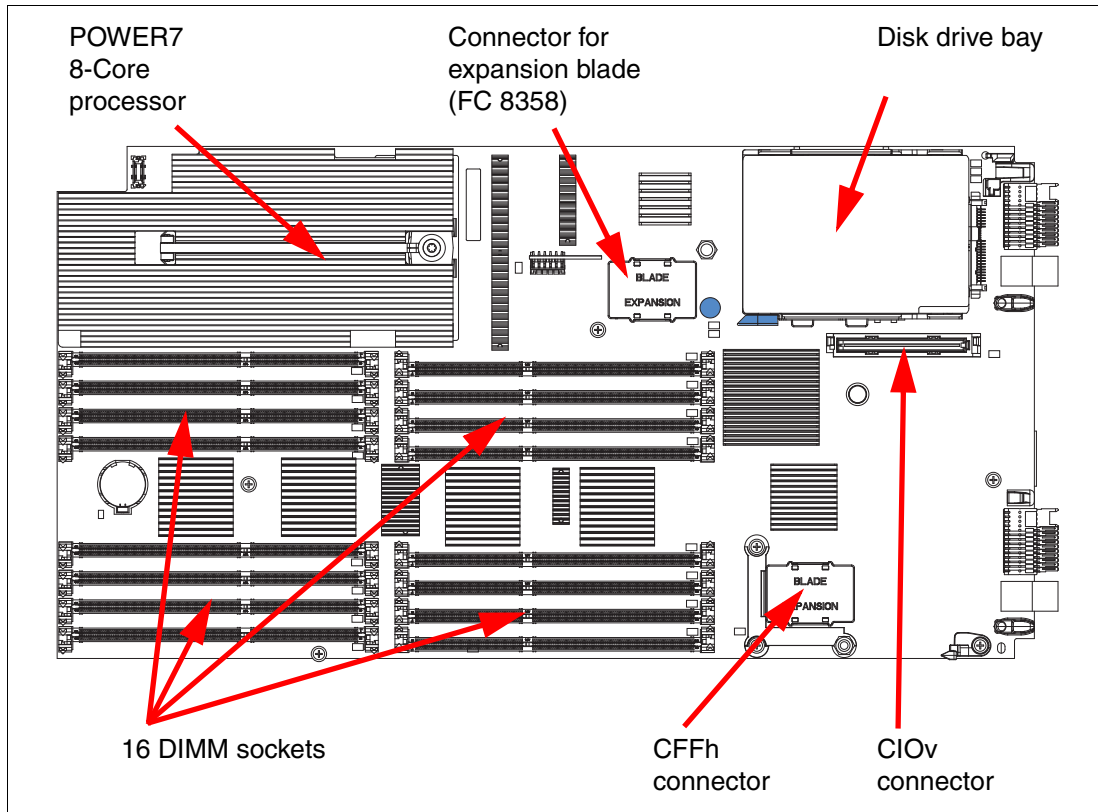


Figure 1-11 Top view of the PS701 blade server

The features of the server are as follows:

- ▶ Machine type and model number
8406-71Y
- ▶ Form factor
Single-wide (30 mm) blade
- ▶ Processors:
 - Single-socket 8-core 64-bit POWER7 processor operating at a 3.0 GHz clock speed
 - Based on CMOS 12S 45 nm SOI (silicon-on-insulator) technology
 - Power consumption is 150w/socket
 - Single-wide (SW) Blade package
- ▶ Memory
 - 16 DIMM slots
 - Minimum capacity 8 GB, maximum capacity 128 GB
 - Industry standard VLP DDR3 DIMMs
- ▶ Disk
 - One disk drive bays supports one SAS HDD (hard disk drive).

- ▶ On-board integrated features:
 - Service processor (SP)
 - Two 1 GB Ethernet ports (HEA)
 - SAS Controller
 - USB Controller which routes to the USB 2.0 port on the media tray.
 - 1 Serial over LAN (SOL) Console through SP
- ▶ Expansion Card I/O Options:
 - One CIOv expansion card slot (PCIe)
 - One CFFh expansion card slot (PCIe)

1.5.3 PS702 system features

The two halves of the BladeCenter PS702 are shown in Figure 1-12.

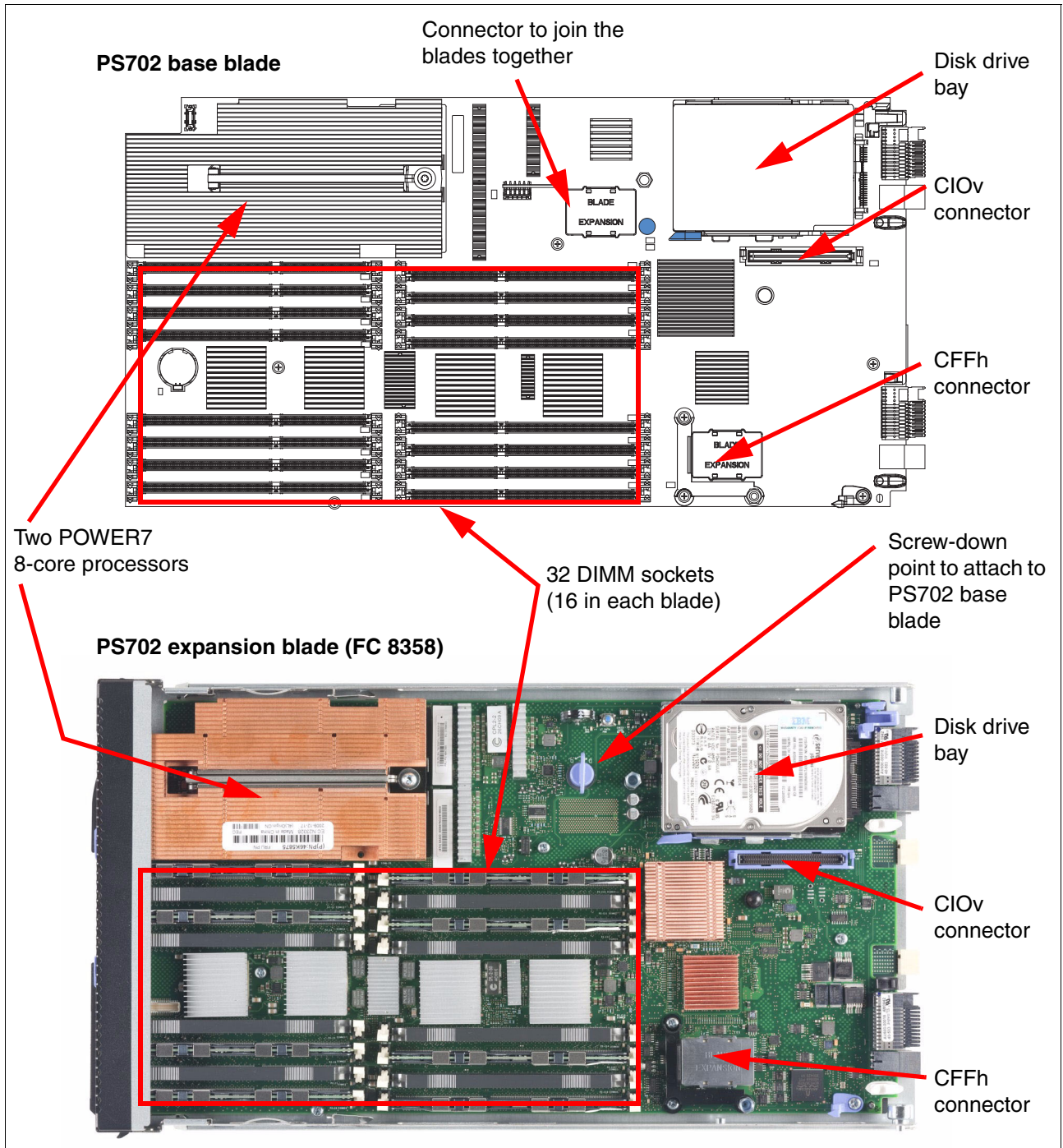


Figure 1-12 Top view of PS702 blade server

The features of the server are as follows:

- ▶ Machine type and model number
 - 8406-71Y + FC 8358
- ▶ Form factor
 - Double-wide (60 mm) blade
- ▶ Processors:
 - Two-socket 16-core 64-bit POWER7 processor operating at a 3.0 GHz clock speed
 - Based on CMOS 12S 45 nm SOI (silicon-on-insulator) technology
 - Power consumption is 150w/socket
 - Double-wide (SW) Blade package
- ▶ Memory
 - 16 DIMM Slots
 - Minimum capacity 8 GB, maximum capacity 256 GB
 - Industry Standard VLP DDR3 DIMMs
- ▶ Disk
 - Two disk drive bays (one on each blade) supports one or two SAS HDD
 - Hardware mirroring RAID 0, RAID 1, RAID 10
- ▶ On-board integrated features:
 - Service processor (one on each blade¹)
 - Four 1 GB Ethernet ports (HEA)
 - SAS Controller
 - USB Controller which routes to the USB 2.0 port on the media tray
 - 1 Serial over LAN (SOL) Console through FSP
- ▶ Expansion Card I/O Options:
 - One CIOv expansion card slot (PCIe)
 - One CFFh expansion card slot (PCIe)

Note: The PS702 is 16-core POWER7 processor-based blade server that is a combination of a single-socket 8-core blade, model 8406-71Y (PS701), and a PS702 expansion blade, feature code #8358.

1.5.4 Minimum features for the POWER7 processor-based blade servers

At the minimum PS700, PS701, and PS702 requires a BladeCenter chassis and one processor socket per blade (four core single socket in PS700, eight core single socket in PS701 and two eight core single socket in PS702 blade servers), minimum memory (8 GB) and zero or one DASD, and a Language Group Specify (mandatory to order voltage nomenclature/language).

Each system has a minimum feature set to be valid. The minimum system configuration for a PS700 or PS701 is shown in Table 1-6 on page 22.

¹ The service processor (or flexible service processor) on the expansion unit provides control but does not offer redundancy with the SP on the base unit.

Table 1-6 Minimum features for PS700 PS701 and PS702 Blade Server

| Category | Minimum features required |
|--|---|
| BladeCenter chassis | Supported BladeCenter chassis (refer to 1.2.1, "Supported BladeCenter chassis" on page 4) |
| Processor | <ul style="list-style-type: none"> ▶ 4-core 3.0 GHz PS700 Blade (#8406-70Y) ▶ 8-core 3.0 GHz PS701 Blade (#8406-71Y) ▶ 16-core 3.0 GHz PS702 Blade (#8406-71Y + FC 8358) |
| Memory | Two DDR3 Memory DIMM: <ul style="list-style-type: none"> ▶ For P700 and P701 8 GB (2 x 4 GB) Memory DIMMs, 1066 MHz (#8208) ▶ For P702 8 GB (2 x 4 GB) Memory DIMMs, 1066 MHz (#8208) on each base board |
| Storage For AIX and Linux: 1x disk drive For IBM i, 2x disk drives | AIX/Linux/Virtual I/O Server: <ul style="list-style-type: none"> ▶ 300 GB SAS 2.5-inch HDD (#8274) OR ▶ 600 GB SAS 2.5 inch HDD (#8276) IBM i (Required VIOS partition) <ul style="list-style-type: none"> ▶ 300 GB SAS 2.5-inch HDD (#8274) OR (▶ 600 GB SAS 2.5 inch HDD (#8276) If Boot from SAN 8 GB Fibre Channel HBA is selected with FC #8240, #8242 or #8271 or Fibre Channel over Ethernet Adapter FC #8275 must be ordered |
| 1x Language Group | Country specific (selected by the customer) |
| Operating system | 1x primary operating system (one of these) <ul style="list-style-type: none"> ▶ AIX (#2146) ▶ Linux (#2147) ▶ IBM i (#2145) plus IBM i 6.1.1 (#0566) |

1.5.5 Power supply features

The power consumption for each PS700, PS701, and PS702 blade server is 12V at 350 watts maximum, which is provided by the BladeCenter power supply modules. The maximum measured value is the worst case power consumption expected from a fully populated server under intensive workload. The maximum measured value also accounts for component tolerance and non-ideal operating conditions. Power consumption and heat load vary greatly by server configuration and use.

Use the IBM Systems Energy Estimator to obtain a heat output estimate based on a specific configuration. The Estimator is available from the following Web page:

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

For information about power supply requirements for each of the BladeCenter chassis supported by POWER7 blade servers and the number of POWER7 blades supported, see 1.2.2, "Number of PS700, PS701, and PS702 blades in a chassis" on page 14.

1.5.6 Processor

The POWER7 3.0 GHz 64-bit POWER7 processor for blade servers is available in four-core (PS700), eight-core (PS701) or two eight-core (PS702) configurations. They are optimized to achieve maximum performance for both the system and its virtual machines. Couple that performance with PowerVM™ and you are now enabled for massive workload consolidation to drive maximum system use, predictable performance, and cost efficiency.

POWER7 Intelligent Threads Technology enables workload optimization by selecting the most suitable threading mode (Single thread (per core) or Simultaneous Multi-thread 2 or 4 modes also called 2-SMT and 4-SMT). The Intelligent Threads Technology can provide improved application performance. In addition, POWER7 processors can maximize cache access to cores, improving performance, using Intelligent Cache technology

POWER7 offers Intelligent Energy Management features that can dramatically and dynamically conserve power and further improve energy efficiency. These features enable the POWER7 processor to operate at a higher frequency if environmental conditions permit, for increased performance and performance per watt. Alternatively, if user settings permit, these features allow the processor to operate at a reduced frequency for significant energy savings.

The key processor feature on each of the P700 blade server are as follows:

- ▶ The PS700 blade server contains one four-core, 64-bit POWER7 3.0 GHz processor with 256 KB per processor core L2 cache and 4 MB per processor core L3 cache. No processor options are available.
- ▶ The PS701 blade server contains one eight-core, 64-bit POWER7 3.0 GHz processor with 256 KB per processor core L2 cache and 4 MB per processor core L3 cache. No processor options are available.
- ▶ The PS702 blade server is a double-wide that supports two eight-core, 64-bit POWER7 3.0 GHz processor with 256 KB per processor core L2 cache and 4 MB per processor core L3 cache. No processor options are available.

1.5.7 Memory features

The PS700, PS701, and PS702 blade servers uses industry standard VLP DDR3 memory DIMMs. Memory DIMMs must be installed in matched pairs with the same size and speed. For details about memory subsystem and layout, see 2.4, “Memory subsystem” on page 46.

The PS700, PS701, and PS702 blade serves have eight, 16, and 32 DIMM slots respectively. Memory is available in 4 GB or 8 GB DIMMs, both operating at a memory speed of 1066 MHz. The memory sizes can be mixed within a system. You can use pairs of 4 GB DIMMs with pairs of 8 GB DIMMs.

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. For details, see 2.4, “Memory subsystem” on page 46. Table 1-7 shows the DIMM features.

Table 1-7 Memory DIMM options

| Feature code | DIMM size | Quantity | Speed |
|--------------|-----------|----------|----------|
| 8208 | 4 GB | 2 | 1066 MHz |
| 8209 | 8 GB | 2 | 1066 MHz |

Notes:

The DDR2 DIMMs used in JS23 and JS43 blade servers are not supported in the Power7 blade servers.

The announcement letter for the POWER7 processor-based blades incorrectly lists the memory speed of the 8 GB DIMMs to be 800 MHz.

1.5.8 I/O features

The PS700 and PS701 have one CIOv PCIe expansion card slot and one CFFh PCIe high-speed expansion card slot. The PS702 blade server has two CIOv expansion card slots and two CFFh expansion card slots.

Table 1-8 shows the supported CIOv and CFFh expansion cards in POWER7 processor-based servers.

Table 1-8 Supported I/O Expansion Card on POWER7 Blades

| Card Description | Feature Code |
|--|--------------|
| CIOv | |
| QLogic 8 Gb Fibre Channel Expansion Card (CIOv) | 8242 |
| QLogic 4 Gb Fibre Channel Expansion Card (CIOv) | 8241 |
| Emulex 8 Gb Fibre Channel Expansion Card (CIOv) | 8240 |
| 3 GB SAS Passthrough Expansion Card (CIOv) | 8246 |
| CFFh | |
| QLogic 8 GB Fibre Channel Expansion Card (CFFh) | 8271 |
| QLogic Ethernet and 4 Gb Fibre Channel Expansion Card (CFFh) | 8252 |
| QLogic 2-port 10 Gb Converged Network Adapter (CFFh) | 8275 |
| 4X InfiniBand DDR Expansion Card (CFFh) | 8258 |

QLogic 8 Gb Fibre Channel Expansion Card (CIOv)

The QLogic 8 Gb Fibre Channel Expansion Card (CIOv) for IBM BladeCenter, feature #8242, enables high-speed access for IBM blade servers to connect to a Fibre Channel storage area network (SAN). When compared to the previous-generation 4 Gb adapters, the new adapter doubles the throughput speeds for Fibre Channel traffic. As a result, you can manage increased amounts of data and possibly benefit from a reduced hardware expense.

The card has the following features:

- ▶ CIOv form factor
- ▶ QLogic 2532 8 Gb ASIC
- ▶ PCI Express 2.0 host interface
- ▶ Support for two full-duplex Fibre Channel ports at 8 Gbps maximum per channel
- ▶ Support for Fibre Channel Protocol Small Computer System Interface (FCP-SCSI) and Fibre Channel Internet Protocol (FC-IP)
- ▶ Support for Fibre Channel service (class 3)
- ▶ Support for switched fabric, point-to-point, and Fibre Channel Arbitrated Loop (FC-AL) connections
- ▶ Support for NPIV when installed in the PS700, PS701, and PS702 Blade Servers

For more information, see the IBM Redbooks at-a-glance guide at the following Web page:

<http://www.redbooks.ibm.com/abstracts/tips0692.html?Open>

QLogic 4 Gb Fibre Channel Expansion Card (CIOv)

The QLogic 4 Gb Fibre Channel Expansion Card (CIOv) for BladeCenter, feature #8241, enables you to connect the BladeCenter servers with CIOv expansion slots to a Fibre Channel SAN. Pick any Fibre Channel storage solution from the IBM System Storage® DS3000, DS4000®, DS5000, and DS8000® series, and begin accessing data over a high-speed interconnect. This card is installed into the PCI Express CIOv slot of a supported blade server. It provides connections to Fibre Channel-compatible modules located in bays 3 and 4 of a supported BladeCenter chassis. A maximum of one QLogic 4 Gb Fibre Channel Expansion Card (CIOv) is supported per single-wide (30 mm) blade server.

The card has the following features:

- ▶ CIOv form factor
- ▶ PCI Express 2.0 host interface
- ▶ Support for two full-duplex Fibre Channel ports at 4 Gbps maximum per channel
- ▶ Support for Fibre Channel Protocol SCSI (FCP-SCSI) and Fibre Channel Internet Protocol (FC-IP)
- ▶ Support for Fibre Channel service (class 3)
- ▶ Support for switched fabric, point-to-point, and Fibre Channel Arbitrated Loop (FC-AL) connections

For more information, see the IBM Redbooks at-a-glance guide at the following Web page:

<http://www.redbooks.ibm.com/abstracts/tips0695.html?Open>

Emulex 8 Gb Fibre Channel Expansion Card (CIOv)

The Emulex 8 Gb Fibre Channel Expansion Card (CIOv) for IBM BladeCenter, feature #8240, enables high-performance connection to a SAN. The innovative design of the IBM BladeCenter midplane enables this Fibre Channel adapter to operate without the need for an optical transceiver module. This saves significant hardware costs. Each adapter provides dual paths to the SAN switches to ensure full redundancy. The exclusive firmware-based architecture allows firmware and features to be upgraded without taking the server offline or rebooting and without the need to upgrade the driver.

The card has the following features:

- ▶ Support of the 8 Gbps Fibre Channel standard
- ▶ Use of the Emulex "Saturn" 8 Gb Fibre Channel I/O Controller (IOC) chip
- ▶ Enablement of high-speed and dual-port connection to a Fibre Channel SAN
- ▶ Can be combined with a CFFh card on the same blade server
- ▶ Comprehensive virtualization capabilities with support for N_Port ID Virtualization (NPIV) and Virtual Fabric
- ▶ Simplified installation and configuration using common HBA drivers
- ▶ Efficient administration by using HBAnyware for HBAs anywhere in the SAN
- ▶ Common driver model that eases management and enables upgrades independent of HBA firmware
- ▶ Support of BladeCenter Open Fabric Manager
- ▶ Support for NPIV when installed in the PS700, PS701, and PS702 blade servers

For more information, see the IBM Redbooks at-a-glance guide at the following Web page:

<http://www.redbooks.ibm.com/abstracts/tips0703.html?Open>

3 Gb SAS Passthrough Expansion Card (CIOv)

This card, feature #8246, is an expansion card that offers the ideal way to connect the supported BladeCenter servers to a wide variety of SAS storage devices. The SAS connectivity card can connect to the Disk Storage Modules in the BladeCenter S. The card routes the pair of SAS channels from the blade's onboard SAS controller to the SAS switches installed in the BladeCenter chassis.

Tip: This card is also known as the SAS Connectivity Card (CIOv) for IBM BladeCenter.

This card is installed into the CIOv slot of the supported blade server. It provides connections to SAS modules located in bays 3 and 4 of a supported BladeCenter chassis.

The card has the following features:

- ▶ CIOv form factor
- ▶ Provides external connections for the two SAS ports of blade server's onboard SAS controller
- ▶ Support for two full-duplex SAS ports at 3 Gbps maximum per channel
- ▶ Support for SAS, SSP, and SMP protocols
- ▶ Connectivity to SAS storage devices

For more information, see the IBM Redbooks at-a-glance guide at the following Web page:

<http://www.redbooks.ibm.com/abstracts/tips0701.html?open>

QLogic 8 Gb Fibre Channel Expansion Card (CFFh)

The QLogic 8 Gb Fibre Channel Expansion Card (CFFh) for IBM BladeCenter, feature #8271, is installed in the blade server and allows connectivity to high-speed switch bays. This expansion card provides flexibility for connecting the blade server to the horizontally oriented BladeCenter H modules in bays 7 and 8 or bays 9 and 10 when using the Multi-Switch Interconnect Module (MSIM). This card is used in conjunction with MSIM on the chassis and requires that a Fibre Channel capable I/O module is installed in the right position of the MSIM. It can be combined with a CFFv I/O card on the same high-speed blade server.

The card has the following features:

- ▶ Support for Fibre Channel protocol SCSI (FCP-SCSI) and Fibre Channel Internet protocol (FCP-IP)
- ▶ Support for point-to-point fabric connection (F-port fabric login)
- ▶ Support for Fibre Channel service (classes 2 and 3)
- ▶ Support for NPIV when installed in PS701, PS700 and PS702 Blade Servers
- ▶ Support for remote startup (boot) operations
- ▶ Support for BladeCenter Open Fabric Manager
- ▶ Support for Fibre Device Management Interface (FDMI) standard (VESA standard)
- ▶ Fibre Channel 8 Gbps, 4 Gbps, or 2 Gbps auto-negotiation

Note: This card is also known as the QLogic Ethernet and 8 Gb Fibre Channel Expansion Card (CFFh) for IBM BladeCenter. However, the Ethernet ports were not supported in the POWER7 processor-based blades at the time of writing.

For more information, see the IBM Redbooks at-a-glance guide at the following Web page:

<http://www.redbooks.ibm.com/abstracts/tips0690.html?Open>

QLogic Ethernet and 4 Gb Fibre Channel Expansion Card (CFFh)

The QLogic Ethernet and 4 Gb Fibre Channel Expansion Card, feature #8252, is a CFFh high speed blade server expansion card with two 4 Gb Fibre Channel ports and two 1 Gb Ethernet ports. It provides QLogic 2432M PCI-Express x4 ASIC for 4 Gb 2-port Fibre Channel and Broadcom 5715S PCI-Express x4 ASIC for 1 Gb 2-port Ethernet. This card is used in conjunction with the Multi-Switch Interconnect Module and is installed in the left position of the MSIM and a Fibre Channel capable I/O module is installed in the right position of the MSIM. Both switches do not need to be present at the same time because the Fibre Channel and Ethernet networks are separate and distinct. It can be combined with a CFFv I/O card on the same high-speed blade server.

The card has the following features:

- ▶ Support for Fibre Channel protocol SCSI (FCP-SCSI) and Fibre Channel Internet protocol (FCP-IP)
- ▶ Support for point-to-point fabric connection (F-port fabric login). Support for remote startup (boot) operations
- ▶ Support for BladeCenter Open Fabric Manager

For more detail see the IBM Redbooks publication *IBM BladeCenter Products and Technology*, SG24-7523, available at the following Web page:

<http://www.redbooks.ibm.com/abstracts/sg247523.html?Open>

QLogic 2-port 10 Gb Converged Network Adapter (CFFh)

The QLogic 2-port 10 Gb Converged Network Adapter (CFFh) for IBM BladeCenter, feature #8275, offers robust 8 Gb Fibre Channel storage connectivity and 10 Gb networking over a single Converged Enhanced Ethernet (CEE) link. Because this adapter combines the functions of a network interface card and a host bus adapter on a single converged adapter, clients can realize potential benefits in cost, power, and cooling, and data center footprint by deploying less hardware.

The card has the following features:

- ▶ CFFh PCI Express 2.0 x8 adapter
- ▶ Communication module: QLogic ISP8112
- ▶ Support for up to two CEE HSSMs in a BladeCenter H or HT chassis
- ▶ Support for 10 Gb Converged Enhanced Ethernet (CEE)
- ▶ Support for Fibre Channel over Converged Enhanced Ethernet (FCoCEE)
- ▶ Full hardware offload for FCoCEE protocol processing
- ▶ Support for IPv4 and IPv6
- ▶ Support for SAN boot over CEE, PXE boot, and iSCSI boot
- ▶ Support for Wake on LAN

For more information, see the IBM Redbooks at-a-glance guide at the following Web page:

<http://www.redbooks.ibm.com/abstracts/tips0716.html?Open>

4X InfiniBand DDR Expansion Card (CFFh)

The InfiniBand 4X DDR Expansion Card for IBM BladeCenter delivers low-latency and high-bandwidth for performance-driven server and storage clustering applications

The card has the following features:

- ▶ 1.2us MPI ping latency
- ▶ 20 Gbps InfiniBand ports
- ▶ CPU offload of transport operations
- ▶ End-to-end QoS and congestion control
- ▶ Hardware-based I/O virtualization
- ▶ TCP/UDP/IP stateless offload

For more information, see the IBM Redbooks publication *IBM BladeCenter Products and Technology*, SG24-7523, available at the following Web page:

<http://www.redbooks.ibm.com/abstracts/sg247523.html?Open>

1.5.9 Disk features

The PS700 blade servers has two disk bays:

- ▶ In the first bay it can have one 2.5 inch SAS HDD
- ▶ In second bay it can have one 2.5 inch SAS HDD

The PS701 blade servers has one disk bay. In this bay it can have one 2.5 inch SAS HDD.

The PS702 blade servers have two disk bays (one on each of the blade):

- ▶ On the base card it can have one 2.5 inch SAS HDD.
- ▶ On the expansion unit it can have one 2.5 inch SAS HDD.

Table 1-6 lists the supported disk features on the PS700, PS701 and PS701 blade servers

Table 1-9 Supported disk drives

| Feature code | Description |
|--------------|------------------------|
| 8274 | 300 GB 10K SFF SAS HDD |
| 8276 | 600 GB 10K SFF SAS HDD |

1.5.10 Standard onboard features

In this section, we describe the standard on-board features.

Service processor

The service processor (or flexible service processor, FSP) is the main integral part of the blade server. It monitors and manages system hardware, resources, and devices. It does the system initialization, configuration, and thermal/power management. It takes corrective action if required.

The PS700 and PS701 have only one service processor. The PS702 blade server has two FSPs (one on each blade). However, the second service processor is only in IO mode and is not redundant to the one on the base blade.

For more details about service processors, see 2.10, “External disk subsystems” on page 68.

Host Ethernet Adapter (HEA)

The integrated IO Hub provides two 1 GB Ethernet ports also called HEAs (host Ethernet adapters). HEA is part of the Integrated Virtual Ethernet subsystem (IVE). Each HEA has their own MAC address and can have a maximum of 16 logical ports. These logical ports can be

used to communicate to the multiple Lpars, which helps in virtualization/sharing of the Ethernet port without using the Ethernet bridge on the Virtual IO Server.

The PS700 and PS701 blade servers have two 1 GB HEA. The PS702 has four 1 GB HEAs (two on each system board).

For more details about HEA and IVE subsystems, see 2.7, “Integrated Virtual Ethernet” on page 61.

SAS Controller

The integrated SAS controller is used to drive the local SAS storage.

The 3 GB SAS Passthrough expansion card can be used to connect to the BladeCenter SAS switch, which can be connected to the external storage. This SAS passthrough expansion card can also be used to connect to BladeCenter S internal drive SAS drives. See “3 Gb SAS Passthrough Expansion Card (CIOv)” on page 26 for more information.

The PS700 and PS701 blades server has one integrated SAS controller. The PS702 has two integrated SAS controllers, one on each blade. The SAS controller host PCI-X interface to P5IOC2 I/O Hub is 64 bits wide and operates at 133 MHz.

The integrated SAS controller supports hardware mirroring RAID 0, RAID 1 or RAID 10 when two HDDs are used in P701 or P702 blade servers

For more information, see “SAS adapter” on page 58 and 2.9, “Internal storage” on page 65

USB controller

The USB controller connects the USB bus to the midplane, which is then routed to the media tray in the BladeCenter chassis to connect to USB devices (such as an optical drive or diskette drive).

For more information, see 2.6.6, “Embedded USB controller” on page 60.

Serial over LAN (SOL)

The integrated SOL function routes the console data stream over standard dual 1 GB Ethernet ports to the Advance Management Module. The PS700, PS701, and PS702 do not have on-board video chips and do not support KVM connections. Console access is only by SOL connection. Each blade can have a single SOL session, however there can be multiple telnet or ssh sessions to the BladeCenter AMM each acting as a SOL connection to a different blade.

For more information, see 2.8.1, “Server console access by SOL” on page 63.

1.6 Supported BladeCenter I/O modules

With IBM BladeCenter, the switches and other I/O modules are installed in the chassis rather than as discrete devices installed in the rack.

The BladeCenter chassis supports a wide variety and range of I/O switch modules. These switch modules are matched to the type, slot location, and form factor of the expansion cards installed in a blade server. For more information, see 1.5.8, “I/O features” on page 24 and 2.6, “Internal I/O subsystem” on page 52.

The I/O switch modules described in the following sections are matched with the on-board HEAs and supported expansion cards in the PS700, PS701, and PS702 blades. In general, the integrated ports on the blades and the additional ports on the expansion cards can function with a single supporting I/O switch module. However, I/O switch modules should be added in pairs to eliminate single points of failure.

For the latest and most current information about blade, expansion card, switch module, and chassis compatibility and interoperability see the *IBM BladeCenter Interoperability Guide* at the following Web page:

<http://www.ibm.com/support/docview.wss?uid=psg1MIGR-5073016>

1.6.1 Ethernet switch and intelligent pass through modules

Various types of Ethernet switch and pass through modules from several manufacturers are available for BladeCenter, and they support different network layers and services. These I/O modules provide external and chassis blade-to-blade connectivity.

The HEAs are on-blade ports that are part of the IVE subsystem that is a standard part of the PS700, PS701, and PS702 blades. For more information, see 2.7, “Integrated Virtual Ethernet” on page 61. There are two physical ports on the PS700 and PS701 and four physical ports on the PS702. The data traffic from these on-blade 1 Gb Ethernet adapters is directed to I/O switch bays 1 and 2 on all BladeCenter chassis except BladeCenter S. On the BladeCenter S the connections for all blade HEA ports are wired to I/O switch bay 1.

To provide external network connectivity and a SOL system console through the BladeCenter Advanced Management Module, at least one Ethernet I/O module is required in switch bay 1. For more information, see 2.8.1, “Server console access by SOL” on page 63.

In addition to the HEA ports, the QLogic Ethernet and 4 Gb Fibre Channel Expansion Card (CFFh) card can provide two additional 1 Gb Ethernet ports per card.

A list of available Ethernet I/O modules that support the on-blade HEA ports and expansion card are shown in Table 1-10 on page 30. Not all switches are supported in every configuration of BladeCenter. Complete compatibility matrixes are available on the following Web pages:

- ▶ ServerProven®:
 - <http://www.ibm.com/servers/eserver/serverproven/compat/us/eserver.html>
- ▶ *BladeCenter Interoperability Guide*
 - <http://www.ibm.com/support/docview.wss?uid=psg1MIGR-5073016>

Table 1-10 Ethernet switch modules

| Part number | Feature Code | Option description | Number x type of external ports | Network layers |
|-------------|--------------|--|---|----------------|
| 43W4395 | 5450 | Cisco Catalyst Switch Module 3012 | 4 x Gigabit Ethernet | Layer 2/3 |
| 41Y8523 | 2989 | Cisco Catalyst Switch Module 3110G | 4 x Gigabit Ethernet, 2 x StackWise Plus | Layer 2/3 |
| 41Y8522 | 2988 | Cisco Catalyst Switch Module 3110X | 1 x 10 Gb Ethernet, 2 x StackWise Plus | Layer 2/3 |
| 39Y9324 | 1484 | IBM Server Connectivity Module | 6 x Gigabit Ethernet | Layer 2 |
| 32R1860 | 1495 | BNT L2/3 Copper Gigabit Ethernet Switch Module | 6 x Gigabit Ethernet | Layer 2/3 |

| Part number | Feature Code | Option description | Number x type of external ports | Network layers |
|-------------|--------------|---|---|----------------|
| 32R1861 | 1496 | BNT L2/3 Fibre Gigabit Ethernet Switch Module | 6 x Gigabit Ethernet | Layer 2/3 |
| 32R1859 | 1494 | BNT Layer 2-7 Gigabit Ethernet Switch Module | 4 x Gigabit Ethernet | Layer 2/7 |
| 44W4404 | 1590 | BNT 1/10 Gb Uplink Ethernet Switch Module | 3 x 10 Gb Ethernet, 6 x Gigabit Ethernet | Layer 2/3 |
| 44W4483 | 5452 | Intelligent Copper Pass-thru Module | 14 x Gigabit Ethernet | - |

1.6.2 SAS I/O modules

SAS I/O modules provide affordable storage connectivity for BladeCenter chassis using SAS technology to create simple fabric for external shared or non-shared storage attachments. A SAS module can also perform RAID controller functions inside the BladeCenter S chassis for HDDs installed into Disk Storage Module (DSM) and external EXP3000 expansions. The SAS RAID Controller Module and DSMs in a BladeCenter S provides RAID 0, 1, 5, and 10 support

In the PS700, PS701, and PS702 blades, the 3 Gb SAS Passthrough Expansion Card (CIOv) is required for external SAS connectivity. The SAS expansion card requires SAS I/O modules in switch bays 3 and 4 of all supported BladeCenters.

Table 1-11 lists the SAS I/O modules and support matrix.

Table 1-11 SAS I/O modules supported by the SAS pass through card

| Part number | Feature code | Description | 3 Gb SAS pass-thru card | BC-E | BC-H | BC-HT | BC-S | MSIM | MSIM-HT |
|-------------|--------------|----------------------------|-------------------------|------|------|-------|------|------|---------|
| 39Y9195 | 2980 | SAS Connectivity Module | Yes | Yes | Yes | Yes | Yes | No | No |
| 43W3584 | 3734 | SAS RAID Controller Module | Yes | No | No | No | Yes | No | No |

1.6.3 Fibre Channel switch and pass-thru modules

Fibre Channel I/O modules are available from several manufacturers. These I/O modules can provide full SAN fabric support up to 8 Gb.

The following 4 Gb and 8 Gb Fibre Channel cards are CIOv form factor and require a Fibre Channel switch or Intelligent Pass Through module in switch bays 3 and 4 of all supported BladeCenters. The CIOv expansion cards are as follows:

- ▶ Emulex 8 Gb Fibre Channel Expansion Card (CIOv)
- ▶ QLogic 4 Gb Fibre Channel Expansion Card (CIOv)
- ▶ QLogic 8 Gb Fibre Channel Expansion Card (CIOv)

Additional 4 Gb and 8 Gb Fibre Channel ports are also available in the CFFh form factor expansion cards. These cards require the use of the MSIM in a BladeCenter H or a MSIM-HT in a BladeCenter HT plus Fibre Channel I/O modules. The CFFh Fibre Channel cards are as follows:

- ▶ QLogic Ethernet and 4 Gb Fibre Channel Expansion Card (CFFh)
- ▶ QLogic 8 Gb Fibre Channel Expansion Card (CFFh)

A list of available Fibre Channel I/O modules that support the CIOv and CFFh expansion cards is shown in Table 1-12. Not all modules are supported in every configuration of BladeCenter. Complete compatibility matrixes are available on the following Web pages:

- ▶ ServerProven:

<http://www.ibm.com/servers/eserver/serverproven/compat/us/eserver.html>

- ▶ *BladeCenter Interoperability Guide*

<http://www.ibm.com/support/docview.wss?uid=psg1MIGR-5073016>

Table 1-12 Fibre Channel I/O modules

| Part number | Feature Code | Description | Number of external ports | Port interface bandwidth |
|-------------|--------------|--|--------------------------|--------------------------|
| 32R1812 | 1569 | Brocade 20-port SAN Switch Module | 6 | 4 Gbps |
| 32R1813 | 1571 | Brocade 10-port SAN Switch Module ^a | 6 | 4 Gbps |
| 42C1828 | 5764 | Brocade Enterprise 20-port 8 Gb SAN Switch Module | 6 | 8 Gbps |
| 44X1920 | 5481 | Brocade 20-port 8 Gb SAN Switch Module | 6 | 8 Gbps |
| 44X1921 | 5483 | Brocade 10-port 8 Gb SAN Switch Module | 6 | 8 Gbps |
| 39Y9280 | 2983 | Cisco Systems 20-port 4 Gb FC Switch Module | 6 | 4 Gbps |
| 39Y9284 | 2984 | Cisco Systems 10-port 4 Gb FC Switch Module ^a | 6 | 4 Gbps |
| 26R0881 | 1560 | QLogic 20-port 4 Gb Fibre Channel Switch Module | 6 | 4 Gbps |
| 43W6725 | 2987 | QLogic 20-port 4 Gb SAN Switch Module | 6 | 4 Gbps |
| 43W6724 | 2986 | QLogic 10-port 4 Gb SAN Switch Module ^a | 6 | 4 Gbps |
| 43W6723 | 2985 | QLogic 4 Gb Intelligent Pass-thru Module ^b | 6 | 4 gbps |
| 44X1905 | 5478 | QLogic 20-Port 8 Gb SAN Switch Module | 6 | 8 Gbps |
| 44X1907 | 5482 | QLogic 8 Gb Intelligent Pass-thru Module ^b | 6 | 8 Gbps |
| 46M6172 | 4799 | QLogic Virtual Fabric Extension Module | 6 | 8 Gbps |

a. Only 10 ports are activated on these switches. An optional upgrade to 20 ports (14 internal + 6 external) is available.

b. Can be upgraded to full fabric switch

1.6.4 Converged networking I/O modules

There are two basic solutions to implement Fibre Channel over Ethernet (FCoE) over a converged network with a BladeCenter.

- ▶ The first solution uses a top-of-rack FCoE capable switch in conjunction with converged network capable 10 Gb Ethernet I/O modules in the BladeCenter. The FCoE-capable top-of-rack switch provides connectivity to the SAN.
- ▶ The second BladeCenter H solution uses a combination of converged network capable 10 Gb Ethernet switch modules and fabric extension modules to provide SAN connectivity, all contained within the BladeCenter H I/O bays.

Implementing either solution with the PS700, PS701, and PS702 blades requires the QLogic 2-port 10 Gb Converged Network Adapter (CFFh). The QLogic Converged Network Adapter (CNA) provides 10 Gb Ethernet and 8 Gb Fibre Channel connectivity over a single CEE link. This card is a CFFh form factor with connections to BladeCenter H and HT I/O module bays 7 and 9.

Table 1-13 shows the currently available I/O modules that are available to provide a FCoE solution.

Table 1-13 Converged network modules supported by the QLogic CNA

| Part Number | Feature Code | Description | Number of external ports |
|-------------|--------------|---|--------------------------|
| 46C7191 | 1639 | BNT Virtual Fabric 10 Gb Switch Module for IBM BladeCenter ^{a b} | 10 x 10 Gb SFP+ |
| 46M6181 | 1641 | 10 Gb Ethernet Pass-Thru Module for BladeCenter ^a | 14 x 10 Gb SFP+ |
| 46M6172 | 4799 | QLogic Virtual Fabric Extension Module for IBM BladeCenter ^{c d} | 6 x 8 Gb FC SFP |
| 46M6071 | 0072 | Cisco Nexus 4001I Switch Module for IBM BladeCenter ^{a e} | 6 x 10 Gb SFP+ |

- a. Used for top-of-rack solution.
- b. Use with Fabric Extension Module for self contain BladeCenter solution.
- c. Also requires BNT Virtual Fabric 10 Gb Switch Module.
- d. BladeCenter H only.
- e. Support is planned.

1.6.5 InfiniBand switch module

The Voltaire 40 Gb InfiniBand Switch Module for BladeCenter provides InfiniBand QDR connectivity between the blade server and external InfiniBand fabrics in non-blocking designs, all on a single device. Voltaire's high speed module also accommodates performance-optimized fabric designs using a single BladeCenter chassis or stacking multiple BladeCenter chassis without requiring an external InfiniBand switch.

The InfiniBand switch module offers 14 internal ports, one to each server, and 16 ports out of the chassis per switch.

The module's HyperScale architecture also provides a unique interswitch link or mesh capability to form highly scalable, cost-effective, and low latency fabrics. Because this switch has 16 uplink ports, they can create a meshed architecture and still have unblocked access to data using the 14 uplink ports. This solution can scale from 14 to 126 nodes and offers latency of less than 200 nanoseconds, allowing applications to operate at maximum efficiency.

The PS700, PS701, and PS702 blades connect to the Voltaire switch through the CFFh form factor 4X InfiniBand DDR Expansion Card (CFFh). The card is only supported in a BladeCenter H and the two ports are connected to high speed I/O switch bays 7/8 and 9/10.

The Voltaire 40 Gb InfiniBand Switch Module for the BladeCenter H is shown in Table 1-14.

Table 1-14 InfiniBand switch module for IBM BladeCenter

| Part number | Feature code | Description | Number of external ports | Type of external ports |
|-------------|--------------|--|--------------------------|------------------------|
| 46M6005 | 0057 | Voltaire 40 Gb InfiniBand Switch Module ^a | 16 | 4X QDR (40 Gbps) |

- a. BladeCenter H only

1.6.6 Multi-switch Interconnect Module

The MSIM is a switch module container that fits in the high speed switch bays (bays 7 and 8 or bays 9 and 10) of the BladeCenter H chassis. Up to two MSIMs can be installed in the BladeCenter H. The MSIM supports most standard switch modules. I/O module. I/O module to MSIM compatibility matrixes can be reviewed at the following Web pages:

- ▶ ServerProven:

<http://www.ibm.com/servers/eserver/serverproven/compat/us/eserver.html>

- ▶ *BladeCenter Interoperability Guide*

<http://www.ibm.com/support/docview.wss?uid=psg1MIGR-5073016>

With PS700, PS701, and PS702 blades, the following expansion cards require a MSIM in a BladeCenter H chassis:

- ▶ QLogic Ethernet and 4 Gb FibreChannel Expansion Card (CFFh)
- ▶ QLogic 8 Gb Fibre Channel Expansion Card (CFFh)

The MSIM is shown in Figure 1-13.

Note: The MSIM comes standard without any I/O modules installed. They need to be ordered separately. In addition, the use of MSIM modules requires that all four power modules be installed in the BladeCenter H chassis.

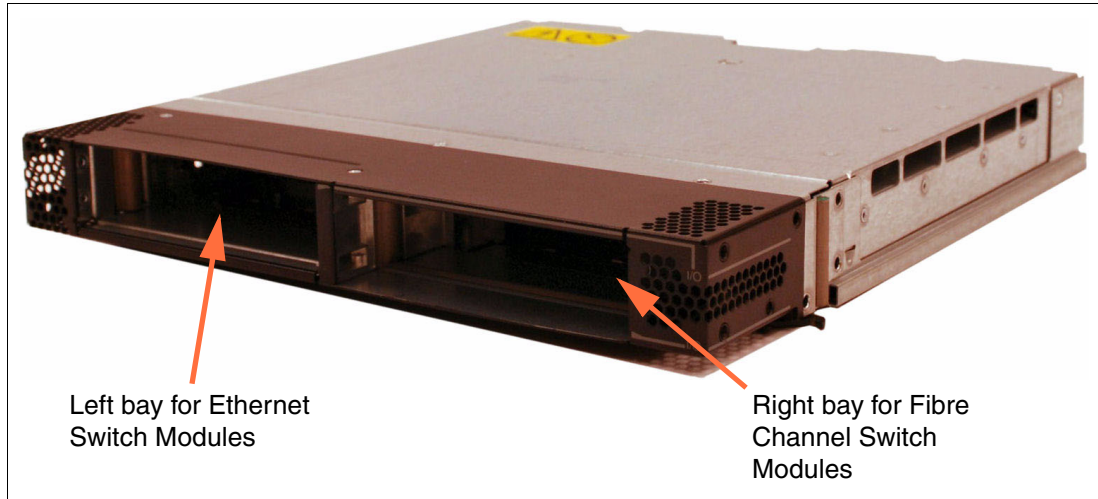


Figure 1-13 Multi-switch Interconnect Module

Table 1-15 shows MSIM ordering information.

Table 1-15 MSIM ordering information

| Description | Part Number | Feature Code |
|--------------------------|-------------|--------------|
| MSIM for IBM BladeCenter | 39Y9314 | 1465 |

1.6.7 Multi-switch Interconnect Module for BladeCenter HT

The Multi-switch Interconnect Module for BladeCenter HT (MSIM-HT) is a switch module container that fits in the high-speed switch bays (bays 7 and 8 or bays 9 and 10) of the BladeCenter HT chassis. Up to two MSIMs can be installed in the BladeCenter HT. The MSIM-HT accepts two supported standard switch modules as shown in Figure 1-14 on page 35.

The MSIM-HT has a reduced number of supported standard I/O modules compared to the MSIM. I/O module to MSIM-HT compatibility matrixes can be viewed at the following Web pages.

For the latest support information, see one of the following resources:

- ▶ ServerProven:
<http://www.ibm.com/servers/eserver/serverproven/compat/us/eserver.html>
- ▶ *BladeCenter Interoperability Guide*
<http://www.ibm.com/support/docview.wss?uid=psg1MIGR-5073016>

With PS700, PS701, and PS702 blades the QLogic Ethernet and 4 Gb FibreChannel Expansion Card (CFFh) requires a MSIM-HT in a BladeCenter HT chassis.

Note: The MSIM-HT comes standard without any I/O modules installed. They need to be ordered separately. In addition, the use of MSIM-HT modules requires that all four power modules be installed in the BladeCenter HT chassis.



Figure 1-14 Multi-switch Interconnect Module for BladeCenter HT

Table 1-16 shows MSIM-HT ordering information.

Table 1-16 MSIM-HT ordering information

| Description | Part Number | Feature Code |
|---|-------------|--------------|
| Multi-switch Interconnect Module for BladeCenter HT | 44R5913 | 5491 |

1.7 Comparison between PS700, PS701, PS702, and 750 models

This section describes the difference between the POWER7 Blade servers and the entry POWER7 Rack Server (Power 750). This helps to better position the POWER7 processor Blade Servers.

The POWER7 Blade Server configuration offers three blade servers. The PS700 is 4-core, the PS701 is 8-Core, and the PS702 is a 16-core Power7 based processor running at 3.0 GHz

The Power 750 offers a a 6-core or 8-core configuration. The 6-core Power7 processor runs at 3.3 GHz and the 8-Core runs at 3.0 GHz, 3.3 GHz or 3.5 GHz systems.

The POWER7 processor has 4 MB L3 cache per core and 256 KB L2 cache per core.

Table 1-17 compares the processor core options and frequencies, L3 cache sizes between the P& Blade servers and entry Power7 rack server Power 750.

Table 1-17 Comparison of P7 Blade Server and P750 Server

| System | Cores per processor | Frequency in (GHz) | L3 cache per processor | Minimum/Maximum cores | Minimum / Maximum memory | Form factor |
|-------------|---------------------|--------------------|------------------------|-----------------------|--------------------------|-------------|
| PS700 blade | 4 | 3.0 | 16 MB | 4 / 4 | 4 GB / 64 GB | Single-wide |
| PS701 blade | 8 | 3.0 | 32 MB | 8 / 8 | 16 GB / 128 GB | Single-wide |
| PS702 blade | 8 | 3.0 | 32 MB | 16 / 16 | 32 GB / 256 GB | Double-wide |
| Power 750 | 6 | 3.3 | 24 MB | 6 / 24 | 8 GB / 512 GB | Rack |
| Power 750 | 8 | 3.0 / 3.3 / 3.55 | 32 MB | 8 / 32 | 8 GB / 512 GB | Rack |

For a detailed comparison, see 2.5, “Technical comparison” on page 51.

1.8 Building to order

You can perform a build to order 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.

The configurator allows you to select an pre-configured Express model or to build a system to order.

The recommendation is to start with one of several 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.

1.9 Model upgrades

The PS700, PS701, and PS702 are new serial-number blade servers. There are no upgrades from POWER5™ or POWER6® blade servers to POWER7 blade servers, which retain the serial number.

However, you can upgrade a PS701 server to a PS702 with the feature code #8358. Feature code 8358 delivers an additional eight-core 3.0 GHz processor, a second set of 16 DIMM slots, and an additional disk bay to the PS701 blade server. Thus you can have an upgrade of two eight-core 3.0 GHz POWER7 processor, with a maximum of 256 GB memory and two 300 GB or 600 GB SAS HDD.



Architecture and technical overview

This chapter discusses the overall system architecture of the POWER7 processor-based blade servers and provides details about each major subsystem and technology.

The topics covered are:

- ▶ 2.1, “Architecture” on page 38
- ▶ 2.2, “The IBM POWER7 processor” on page 38
- ▶ 2.3, “POWER7 processor-based blades” on page 46
- ▶ 2.4, “Memory subsystem” on page 46
- ▶ 2.5, “Technical comparison” on page 51
- ▶ 2.6, “Internal I/O subsystem” on page 52
- ▶ 2.7, “Integrated Virtual Ethernet” on page 61
- ▶ 2.8, “Service processor” on page 63
- ▶ 2.9, “Internal storage” on page 65
- ▶ 2.10, “External disk subsystems” on page 68
- ▶ 2.11, “IVM” on page 71
- ▶ 2.12, “Operating system support” on page 72
- ▶ 2.13, “IBM EnergyScale” on page 74

Note: The bandwidths that are provided throughout the chapter are theoretical maximums used for reference.

2.1 Architecture

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

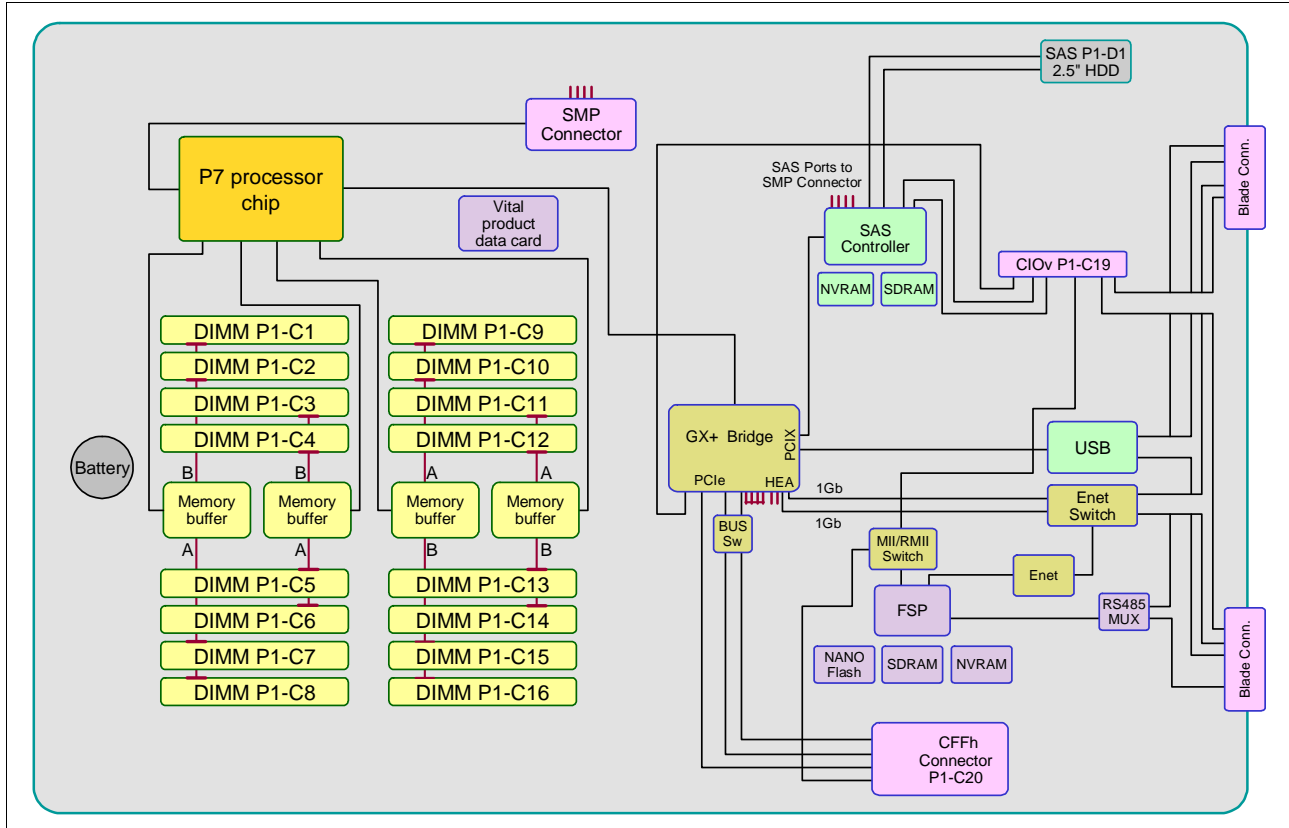


Figure 2-1 PS701 logical data flow

2.2 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 to deliver leading throughput, efficiency, scalability, and reliability, availability, and serviceability (RAS).

Although the processor is an important component in delivering outstanding servers, many elements and facilities have to be balanced across a server 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.

IBM has been innovative 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 elements:

- ▶ 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 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, simultaneous multiprocessing (SMP) links, and memory controllers.

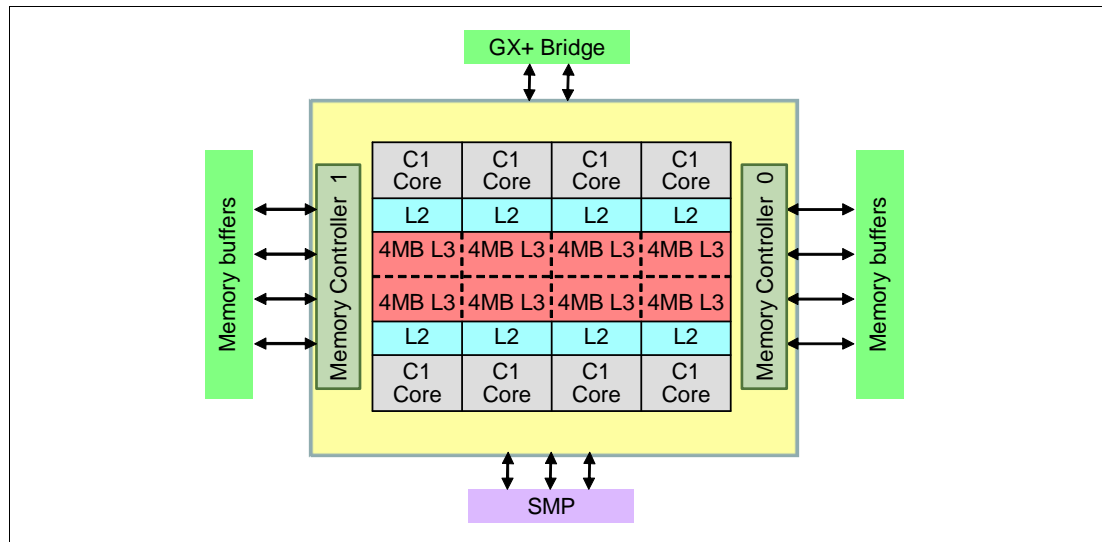


Figure 2-2 POWER7 processor architecture

2.2.1 POWER7 processor overview

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

The POWER7 processor chip is 567 mm² 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 DDR3 (Double Data Rate 3) memory controllers, each with four memory channels. To scale effectively, the POWER7 processor uses a combination of local and global SMP links with high coherency bandwidth and makes use 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 | <ul style="list-style-type: none"> ▶ 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 2.2.6, "On-chip L3 cache innovation and intelligent cache" on page 43) |
| Processor cores | 8 |
| Max execution threads core/chip | 4/32 |
| L2 cache core/chip | 256 KB / 2 MB |
| On-chip L3 cache core/chip | 4 MB / 32 MB |
| DDR3 memory controllers | 2 |
| SMP design-point | Up to 32 sockets with IBM POWER7 processors |
| Compatibility | With prior generation of POWER processor |

2.2.2 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 as follows:

- ▶ 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 caches that are tightly coupled to each POWER7 processor core are as follows:

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

2.2.3 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

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

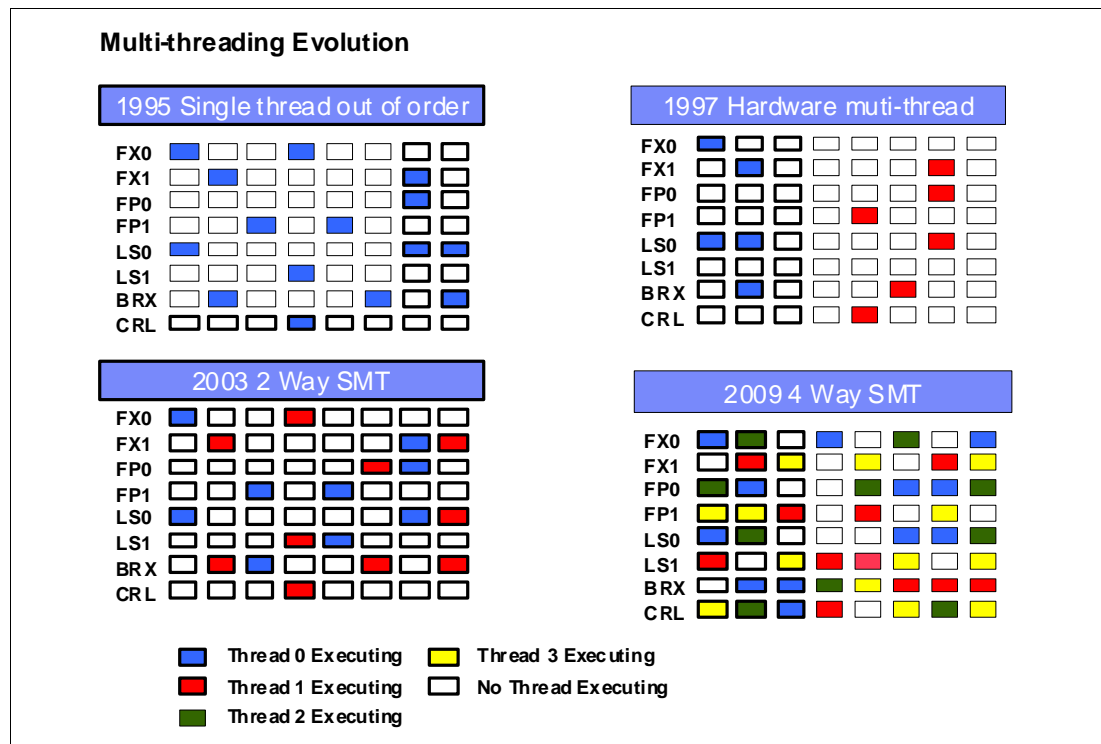


Figure 2-3 Evolution of simultaneous multithreading

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

Intelligent threads

The POWER7 processor features intelligent threads that can vary based on the workload demand. The system either automatically selects (or the system administrator can manually select) whether a workload benefits from dedicating as much capability as possible to a single thread of work, or if the workload benefits more from having capability spread across two or four threads of work. With more threads, the POWER7 processor can deliver more total capacity as more tasks are accomplished in parallel. With fewer threads, workloads that need fast individual tasks can get the performance they need for maximum benefit.

2.2.4 Memory access

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

Note: In certain POWER7 processor-based systems (including the PS700, PS701, and PS702 blades) only one memory controller is active.

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

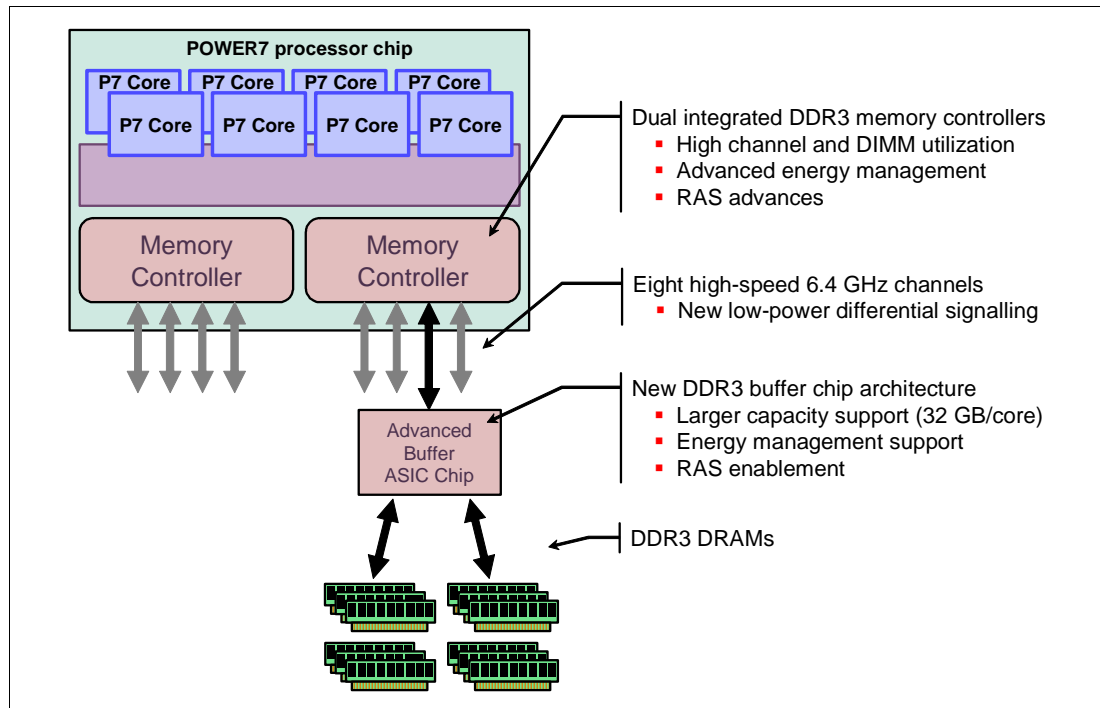


Figure 2-4 Overview of POWER7 memory access structure

2.2.5 Flexible POWER7 processor packaging and offerings

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

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.

The L3 cache associated with the implementation is dependant 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.

Optimized for servers

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 appropriate to the scaling-point of particular server models.

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

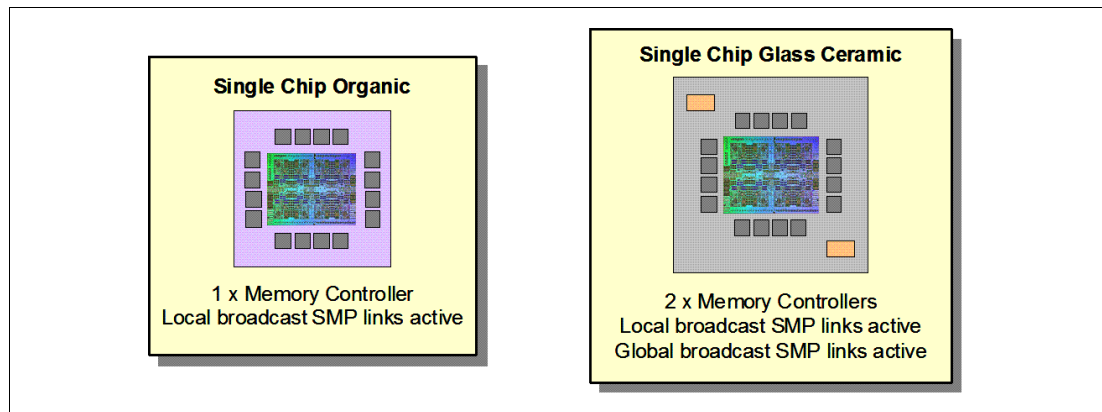


Figure 2-5 Outline of the POWER7 processor physical packaging

2.2.6 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 the cores on the POWER7 processor die.

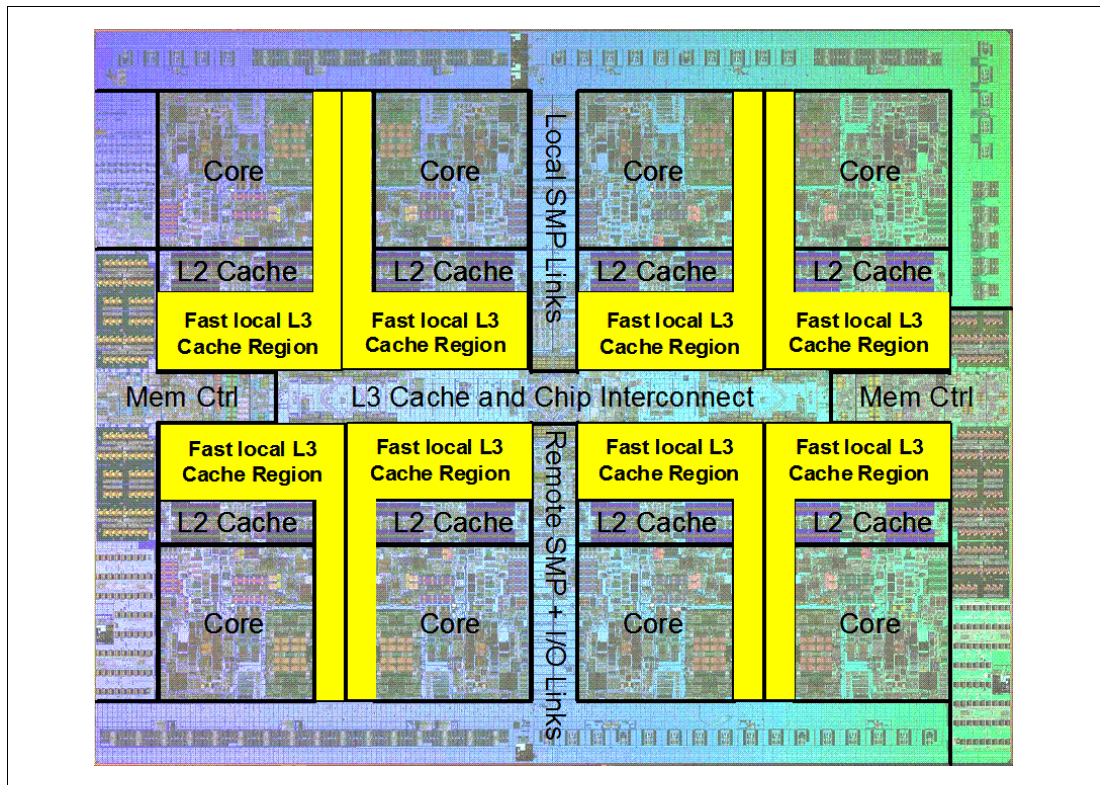


Figure 2-6 FLR-L3 cache regions on the POWER7 processor

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

- ▶ Latency improvement
 - A six-to-one latency improvement occurs by moving the L3 cache on-chip compared to L3 accesses on an external (on-ceramic) ASIC.
- ▶ Bandwidth improvement
 - A 2x bandwidth improvement occurs with on-chip interconnect. Frequency and bus sizes are increased to and from each core.
- ▶ No off-chip driver or receivers
 - Removing drivers and receivers from the L3 access path lowers interface requirements, conserves energy, and lowers latency.
- ▶ Small physical footprint
 - The performance of eDRAM when implemented on-chip is similar to conventional SRAM but requires far less physical space. IBM on-chip eDRAM uses only a third of the components used in conventional SRAM, which has a minimum of six transistors to implement a 1-bit memory cell.
- ▶ Low energy consumption
 - The on-chip eDRAM uses only 20% of the standby power of SRAM.

2.2.7 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 optimize energy usage and performance dynamically, so that the best possible balance is maintained. Intelligent energy features (such as EnergyScale™) work with the BladeCenter Advanced Management Module (AMM) and IBM Systems Director Active Energy Manager™ to optimize processor speed dynamically, based on thermal conditions and system use.

2.2.8 Comparison of the POWER7 and POWER6 processors

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

Note: This shows the characteristics of the POWER7 processors in general, but not necessarily as implemented in the POWER7 processor-based blade servers.

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

| Feature | POWER7 | POWER6+ | POWER6 |
|---------------------------------|---|---------------------------|---------------------------|
| Technology | 45 nm | 65 nm | 65 nm |
| Die size | 567 mm ² | 341 mm ² | 341 mm ² |
| Maximum cores | 8 | 2 | 2 |
| Maximum SMT threads per core | 4 threads | 2 threads | 2 threads |
| Maximum frequency | 4.14 GHz | 5.0 GHz | 4.7 GHz |
| L2 Cache | 256 KB per core | 4 MB 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 | 32 MB off-chip eDRAM ASIC |
| Memory support | DDR3 | DDR2 | DDR2 |
| I/O Bus | Two GX+ | One GX+ | One GX+ |
| Enhanced Cache Mode (TurboCore) | Yes | No | No |
| Sleep & Nap Mode | Both | Nap only | Nap only |

2.3 POWER7 processor-based blades

The PS700 blade contains a single processor socket with a four-core processor and eight DDR3 memory DIMM slots. The PS701 blade contains a single processor socket with an eight-core processor and 16 DDR3 memory DIMM slots. The PS702 blade contains two processor sockets, each with a eight-core processor and a total of 32 DDR3 memory DIMM slots. The cores in all these blades run at 3.0 GHz.

POWER7 processor-based blades support POWER7 processors with various processors core counts. Table 2-3 summarizes the POWER7 processors for the PS700, PS701, and PS702 blades.

Table 2-3 Summary of POWER7 processor options for the PS700, PS701, and PS702 blades

| Blade Model | Cores per POWER7 processor | Number of POWER7 processors | Frequency (GHz) | L3 cache size per POWER7 processor (MB) |
|-------------|----------------------------|-----------------------------|-----------------|---|
| PS700 | 4 | 1 | 3.0 | 16 |
| PS701 | 8 | 1 | 3.0 | 32 |
| PS702 | 8 | 2 | 3.0 | 32 |

2.4 Memory subsystem

The PS700 4-core processor contains one integrated DDR3 memory controller and two memory buffers that can interface with a total of eight DDR3 DIMMS. The PS701 single 8-core processor, and the PS702's two 8-core processors chips also use a single memory controller per processor chip but use four memory buffers that can access a total of 16 or 32 DDR3 DIMMS respectively.

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

2.4.1 Memory placement rules

The minimum DDR3 memory capacity for the PS700, PS701, and PS702 8 GB (2 x 4 GB DIMMs). The maximum memory supported is as follows:

- ▶ PS700 64 GB (8 x 8 GB)
- ▶ PS701 128 GB (16 x 8 GB)
- ▶ PS702 256 GB (32 x 8 GB)

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

Figure 2-7 on page 47 show the PS701 physical memory DIMM topology. Figure 2-8 on page 47 shows the PS701 and PS702 base blade.

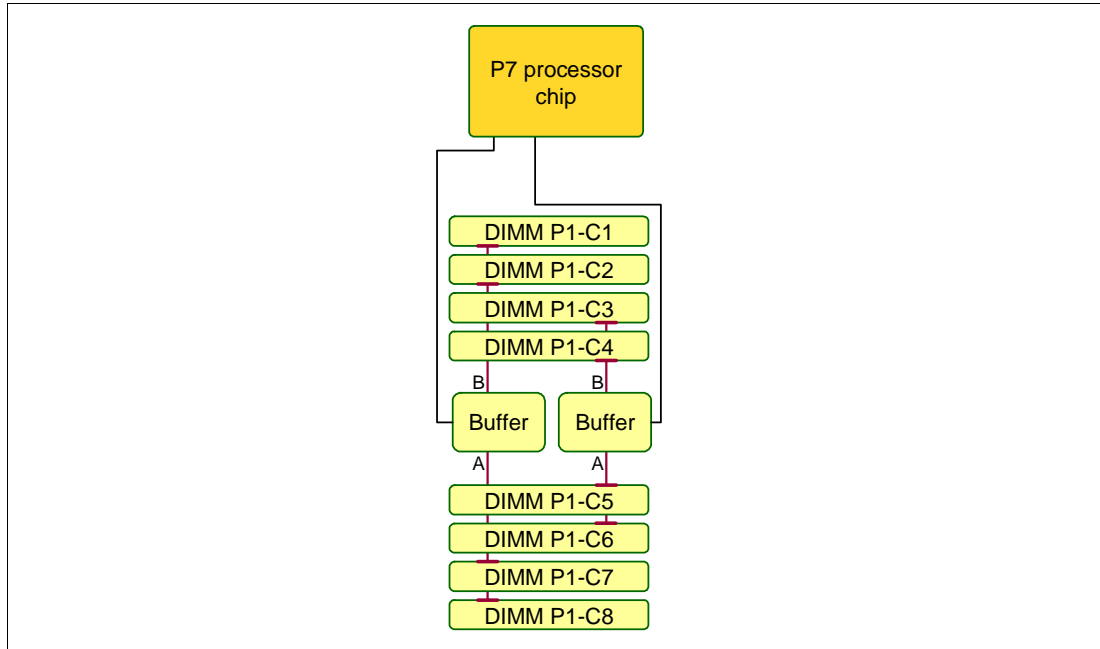


Figure 2-7 Memory DIMM topology for the PS700

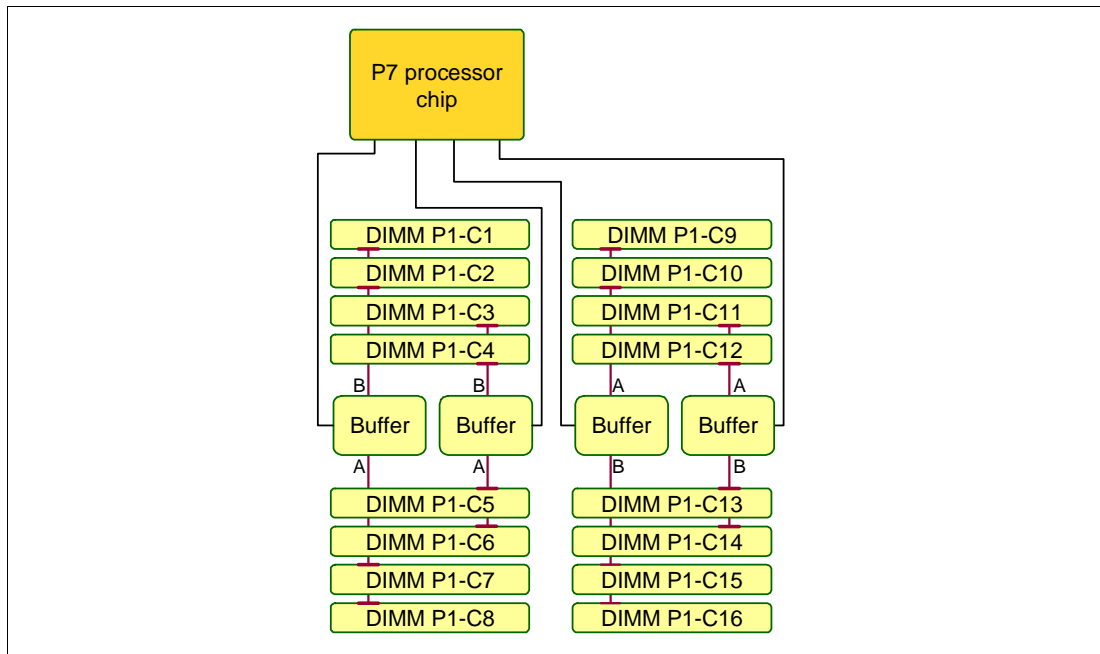


Figure 2-8 Memory DIMM topology for the PS701 and PS702 base blade

There are eight buffered DIMM slots on the PS700, and 16 on the PS701 and PS702 base blade with an additional 16 slots on the PS702 expansion unit. The PS700 DIMM slots are numbered P1-C1 through P1-C8 as shown in Figure 2-7. The PS701 and the PS702 base blade have slots labelled P1-C1 through P1-C16 as shown in Figure 2-8. For the PS702 expansion unit the numbering is the same except for the reference to the second planar board. The numbering is from P2-C1 through P2-C16.

The memory-placement rules are as follows:

- ▶ Memory is installed in DIMM-pairs (as in two DIMMs)
- ▶ DIMM pairs must be matched in size (that is, two 4 GB DIMMs or two 8 GB DIMMs).
- ▶ Minimum memory requirement are as follows:
 - PS700 8 GB (2 x 4 GB DIMMs)
 - PS701 8 GB (2 x 4 GB DIMMs)
 - PS702 8 GB (2 x 4 GB DIMMs)

Note: The stated memory DIMM numbers are the minimums supported by the architecture but might not indicate the minimum order amounts.

- ▶ Mixing of DIMM capacity between pairs is permitted.

DIMMs should be installed in specific DIMM sockets depending on the number of DIMMs to install. This is described in the following three tables.

For the PS700, Table 2-4 shows the required placement of memory DIMMs depending on the number of DIMMs installed (2, 4, 6, or 8).

Table 2-4 PS700 DIMM placement rules

| DIMM socket: | PS700 | | | |
|--------------|-----------------------------|---|---|---|
| | Number of DIMMs to install: | | | |
| | 2 | 4 | 6 | 8 |
| P1-C1 | x | x | x | x |
| P1-C2 | | | x | x |
| P1-C3 | x | x | x | x |
| P1-C4 | | | x | x |
| P1-C5 | | | | x |
| P1-C6 | | x | x | x |
| P1-C7 | | | | x |
| P1-C8 | | x | x | x |

For the PS701, Table 2-5 shows the required placement of memory DIMMs depending on the number of DIMMs installed.

Table 2-5 PS701 DIMM placement rules

| DIMM socket: | PS701 | | | | | | | |
|--------------|-----------------------------|---|---|---|----|----|----|----|
| | Number of DIMMs to install: | | | | | | | |
| | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 |
| P1-C1 | x | x | x | x | x | x | x | x |
| P1-C2 | | | | | x | x | x | x |
| P1-C3 | x | x | x | x | x | x | x | x |
| P1-C4 | | | | | x | x | x | x |
| P1-C5 | | | | | | x | x | x |
| P1-C6 | | | x | x | x | x | x | x |
| P1-C7 | | | | | | x | x | x |
| P1-C8 | | | x | x | x | x | x | x |
| P1-C9 | | | | x | x | x | x | x |
| P1-C10 | | | | | | | | x |
| P1-C11 | | | | x | x | x | x | x |
| P1-C12 | | | | | | | | x |
| P1-C13 | | | | | | | x | x |
| P1-C14 | | x | x | x | x | x | x | x |
| P1-C15 | | | | | | | x | x |
| P1-C16 | | x | x | x | x | x | x | x |

For the PS702, Table 2-6 shows the required placement of memory DIMMs depending on the number of DIMMs installed.

Table 2-6 PS702 DIMM placement rules

| DIMM socket: | PS702 | | | | | | | | | | | | | | | |
|--------------|----------------------------|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|
| | Number of DIMMs to install | | | | | | | | | | | | | | | |
| | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 | 20 | 22 | 24 | 26 | 28 | 30 | 32 |
| P1-C1 | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| P1-C2 | | | | | | | | | x | x | x | x | x | x | x | x |
| P1-C3 | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| P1-C4 | | | | | | | | | x | x | x | x | x | x | x | x |
| P1-C5 | | | | | | | | | | x | x | x | x | x | x | x |
| P1-C6 | | | | | x | x | x | x | x | x | x | x | x | x | x | x |
| P1-C7 | | | | | | | | | | x | x | x | x | x | x | x |
| P1-C8 | | | | | x | x | x | x | x | x | x | x | x | x | x | x |
| P1-C9 | | | | | | | | x | x | x | x | x | x | x | x | x |
| P1-C10 | | | | | | | | | | | | | | | x | x |
| P1-C11 | | | | | | | | x | x | x | x | x | x | x | x | x |
| P1-C12 | | | | | | | | | | | | | | | x | x |
| P1-C13 | | | | | | | | | | | | | x | x | x | x |
| P1-C14 | | | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| P1-C15 | | | | | | | | | | | | | x | x | x | x |
| P1-C16 | | | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| P2-C1 | | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| P2-C2 | | | | | | | | | | x | x | x | x | x | x | x |
| P2-C3 | | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| P2-C4 | | | | | | | | | | x | x | x | x | x | x | x |
| P2-C5 | | | | | | | | | | | | x | x | x | x | x |
| P2-C6 | | | | | | x | x | x | x | x | x | x | x | x | x | x |
| P2-C7 | | | | | | | | | | | | x | x | x | x | x |
| P2-C8 | | | | | | x | x | x | x | x | x | x | x | x | x | x |
| P2-C9 | | | | | | | | | x | x | x | x | x | x | x | x |
| P2-C10 | | | | | | | | | | | | | | | | x |
| P2-C11 | | | | | | | | | | x | x | x | x | x | x | x |
| P2-C12 | | | | | | | | | | | | | | | | x |
| P2-C13 | | | | | | | | | | | | | | | x | x |
| P2-C14 | | | | x | x | x | x | x | x | x | x | x | x | x | x | x |
| P2-C15 | | | | | | | | | | | | | | | x | x |
| P2-C16 | | | | x | x | x | x | x | x | x | x | x | x | x | x | x |

2.5 Technical comparison

Table 2-7 shows a comparison of the technical aspects of the PS700, PS701, and PS702 blade compared to a Power 750 Express.

Table 2-7 Comparison of technical characteristics between PS blades and the Power 750 Express

| Systems characteristic | PS700 | PS701 | PS702 | Power 750 Express |
|---|---------------------|---------------------|---------------------|---|
| Processor | 4-cores at 3.0 GHz | 8-cores at 3.0 GHz | 16-cores at 3.0 GHz | 6-cores at 3.3 GHz 8-cores at 3.0 GHz, 3.3 GHz, 3.55 GHz |
| Pluggable processor cards | Not Applicable | Not Applicable | Not Applicable | 1–4 |
| Min./Max. processor cores | 4 | 8 | 16 | 6/24 (6-core) or 8/32 (8-core) |
| L3 cache | On-chip eDRAM | On-chip eDRAM | On-chip eDRAM | On-chip eDRAM |
| Max memory slots and type | 8 DDR3 | 16 DDR3 | 32 DDR3 | 8 slots per processor card (32 slots max.), DDR3 |
| Memory chipkill | Yes | Yes | Yes | Yes |
| Memory spare | No | No | No | Yes |
| Memory hotplug | No | No | No | No |
| EnergyScale device | Yes | Yes | Yes | Yes |
| PCIe x8 slots | 2 | 2 | 4 | 3 |
| PCI-X 2.0 slots | 0 | 0 | 0 | 2 |
| PCIe and PCI-X hot plug | No | No | No | Yes |
| Integrated Virtual Ethernet Ports / Speed | Integrated 2 / 1 Gb | Integrated 2 / 1 Gb | Integrated 4 / 1 Gb | daughter card quad port / 1 Gb or dual port / 10 Gb |
| PowerVM support | Yes | Yes | Yes | Yes |
| Capacity on Demand | No | No | No | No |
| Redundant hotplug power | Yes through chassis | Yes through chassis | Yes through chassis | Yes |
| DASD bays | 2 | 1 | 2 | 8 (hot-plug, front access, SFF) |
| GX slot (GX+ slot does not support RIO2) | Not applicable | Not applicable | Not applicable | 1 x GX+ slot and 1 x GX++ slot (not hot pluggable) |

2.6 Internal I/O subsystem

Each POWER7 processor as implemented in the POWER7 processor-based blades utilizes a single GX+ bus which is used to connect to the I/O subsystem. The I/O subsystem is a GX+ multifunctional host bridge chip which provides the following major interfaces:

- ▶ One GX+ primary interface to the processor
- ▶ Two 64-bit PCI-X 2.0 buses, one 64-bit PCI-X 1.0 bus, and one 32-bit PCI-X 1.0 bus
- ▶ Four x8 PCI Express links
- ▶ Two 10 Gbps Ethernet ports: Each port is individually configurable to function as two 1 Gbps ports

The PS702 with two POWER7 processors also has two GX+ multifunctional host bridge chips. Unless otherwise noted, references to slots, embedded controllers and so forth are assumed to be doubled for the PS702.

Note: Table 2-2 on page 45 indicates there are two GX+ buses in the POWER7 processor however only one of them is active in the PS700 and PS701, and each processor in the PS702.

2.6.1 Peripheral Component Interconnect Express (PCIe) bus

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 is called a lane. A PCIe link might be comprised of multiple lanes. In such configurations, the connection is labeled as x1, x2, x8, x12, x16, or x32, where the number is the number of lanes.

The PCIe expansion card options for the PS700, PS701, and PS702 blades support Extended Error Handling (EEH). The card ports are routed through the BladeCenter mid-plane to predetermined I/O switch bays. The switches installed in these switch bays must match the type of expansion card installed, Ethernet, Fibre Channel, and so forth.

2.6.2 PCIe slots

The two PCIe slots are connected to the four x8 PCIe links on the GX+ multifunctional host bridge chip. One of the links supports the CIOv connector and the other three links support the CFFh connector on the blade. All PCIe slots are 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 examines the affected bus, allows the device driver to reset it, and continues 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.

Expansion card form factors

There are two PCIe card form factors supported on the PS700, PS701, and PS702 blades:

- ▶ CIOv
- ▶ CFFh

CIOv form factor

A CIOv expansion card uses the PCI Express 2.0 x8 160 pin connector. A CIOv adapter requires compatible switch modules to be installed in bay 3 and bay 4 of the BladeCenter chassis. The CIOv card can be used in any BladeCenter that supports the PS700, PS701, and PS702 blades.

CFFh form factor

The CFFh expansion card attaches to the 450 pin PCIe Express connector of the blade server. In addition, the CFFh adapter can only be used in servers that are installed in the BladeCenter H, BladeCenter HT, or BladeCenter S chassis.

A CFFh adapter requires that either:

- ▶ A Multi-Switch Interconnect Module (MSIM) or MSIM-HT (BladeCenter HT chassis) is installed in bays 7 and 8, bays 9 and 10, or both.
- ▶ A high speed switch module be installed in bay 7 and bay 9.
- ▶ In the BladeCenter S, a compatible switch module is installed in bay 2.

The requirement of either the MSIM, MSIM-HT, or high-speed switch modules depends on the type of CFFh expansion card installed. The MSIM or MSIM-HT must contain compatible switch modules. See 1.6.6, “Multi-switch Interconnect Module” on page 33, or 1.6.7, “Multi-switch Interconnect Module for BladeCenter HT” on page 34, for more information about the MSIM or MSIM-HT.

The CIOv expansion card can be used in conjunction with a CFFh card in BladeCenter H, HT and in certain cases a BladeCenter S chassis, depending on the expansion card type.

Table 2-8 lists the slot types, locations, and supported expansion card form factor types of the PS700, PS701, and PS702 blades.

Table 2-8 Slot configuration of the PS700, PS701, and PS702 blades

| Card location | Form factor | PS700 location | PS701 location | PS702 location |
|----------------------|--------------------|-----------------------|-----------------------|-----------------------|
| Base blade | CIOv | P1-C11 | P1-C19 | P1-C19 |
| Base blade | CFFh | P1-C12 | P1-C20 | P1-C20 |
| Expansion blade | CIOv | Not present | Not present | P2-C19 |
| Expansion blade | CFFh | Not present | Not present | P2-C20 |

Figure 2-9 shows the locations of the PCIe CIOv and CFFh connectors and the physical location codes for the PS700.

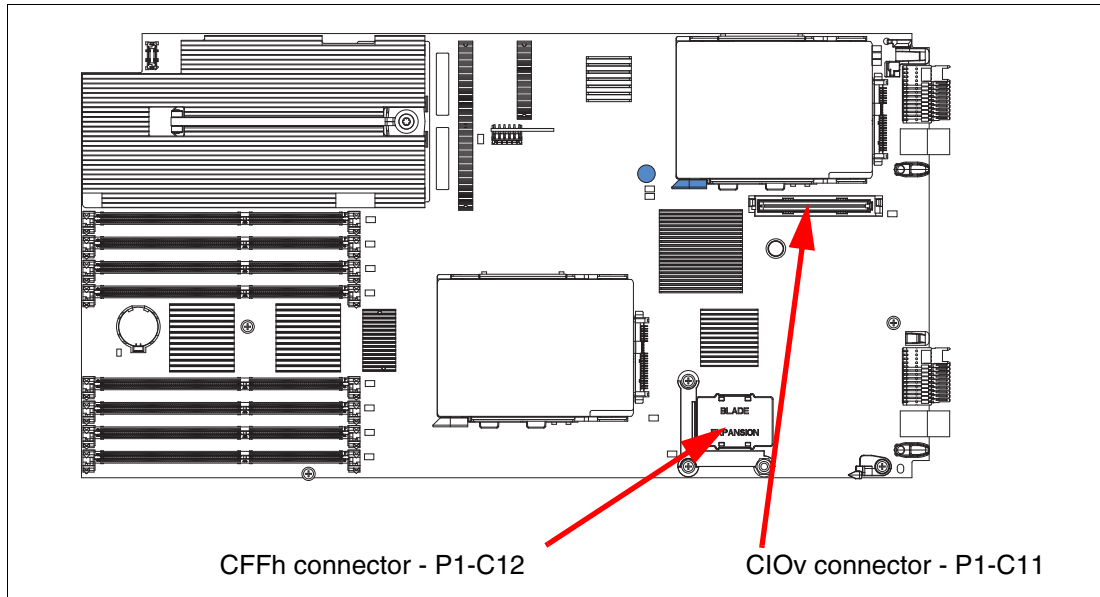


Figure 2-9 PS700 location codes for PCIe expansion cards

Figure 2-10 shows the locations of the PCIe CIOv and CFFh connectors for the PS701 and PS702 base planar and the physical location codes. The expansion unit for the PS702 uses the prefix P2 for the slots on the second planar.

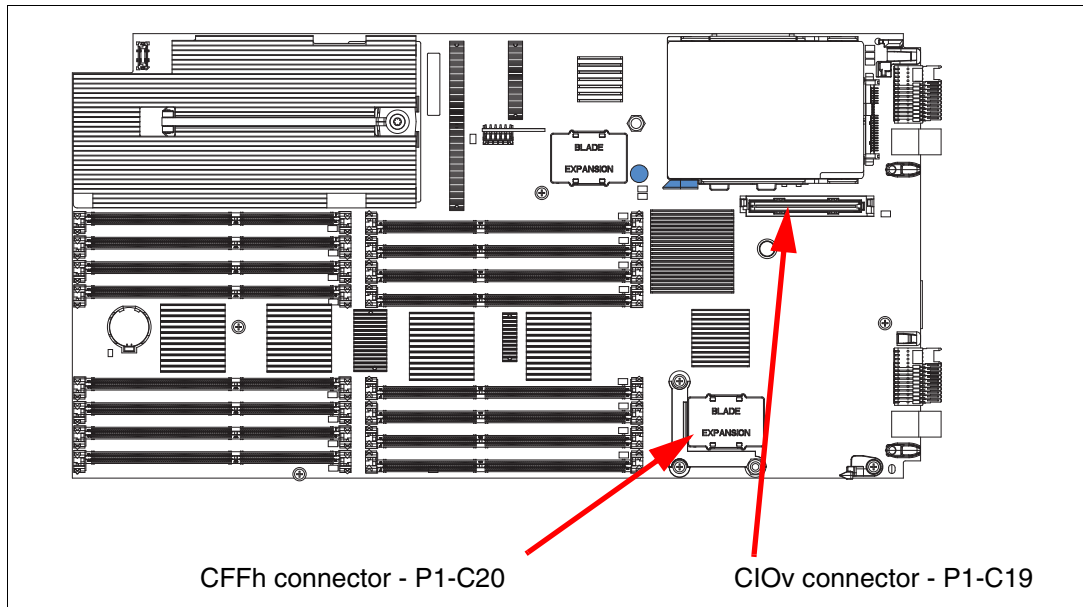


Figure 2-10 PS701 and PS702 base location codes for PCIe expansion cards

Figure 2-11 shows the locations of the PCIe CIOv and CFFh connectors for the PS702 expansion blade (feature code 8358) and the physical location codes.

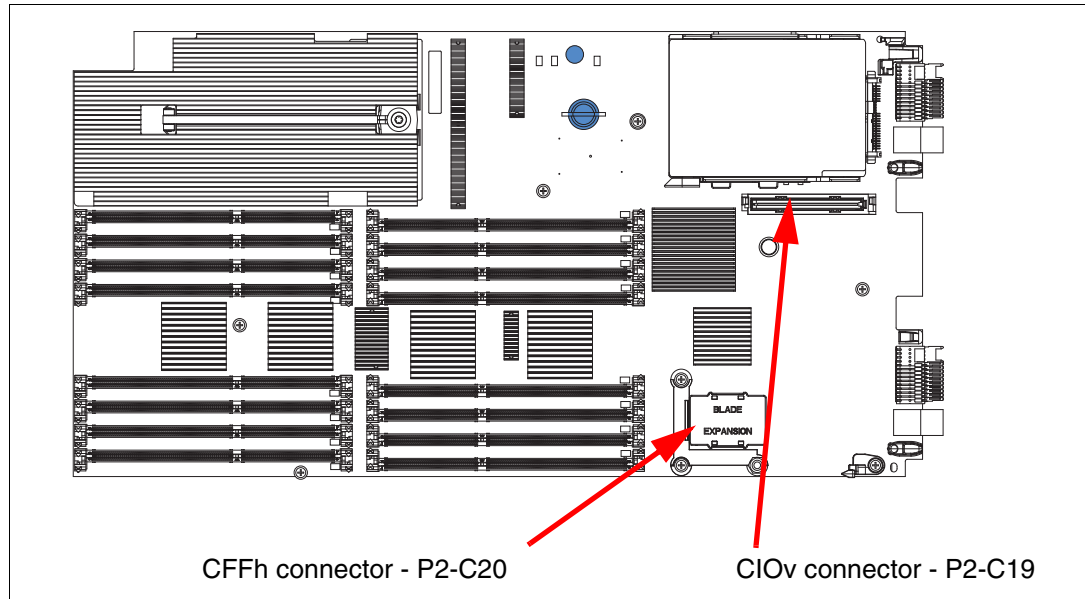


Figure 2-11 PS702 expansion blade location codes for PCIe expansion cards

BladeCenter I/O topology

There are no externally accessible ports on the PS700, PS701, and PS702 blades, All I/O is routed through a BladeCenter midplane to the I/O modules bays.

The I/O ports on all expansion cards are typically set up to provide a redundant pair of ports. Each port has a separate path through the mid-plane of the BladeCenter chassis to a specific I/O module bay. Figure 2-12 on page 56 through Figure 2-15 on page 57 show the four supported BladeCenter chassis and the I/O topology for each.

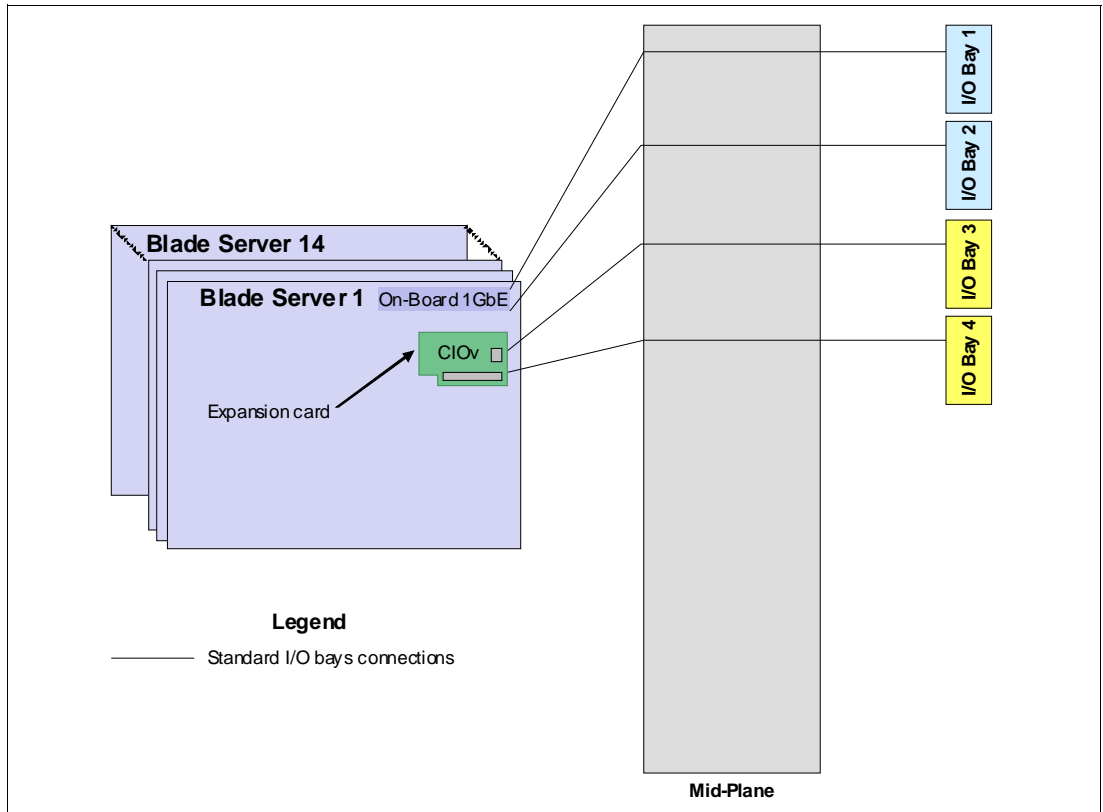


Figure 2-12 BladeCenter E I/O topology

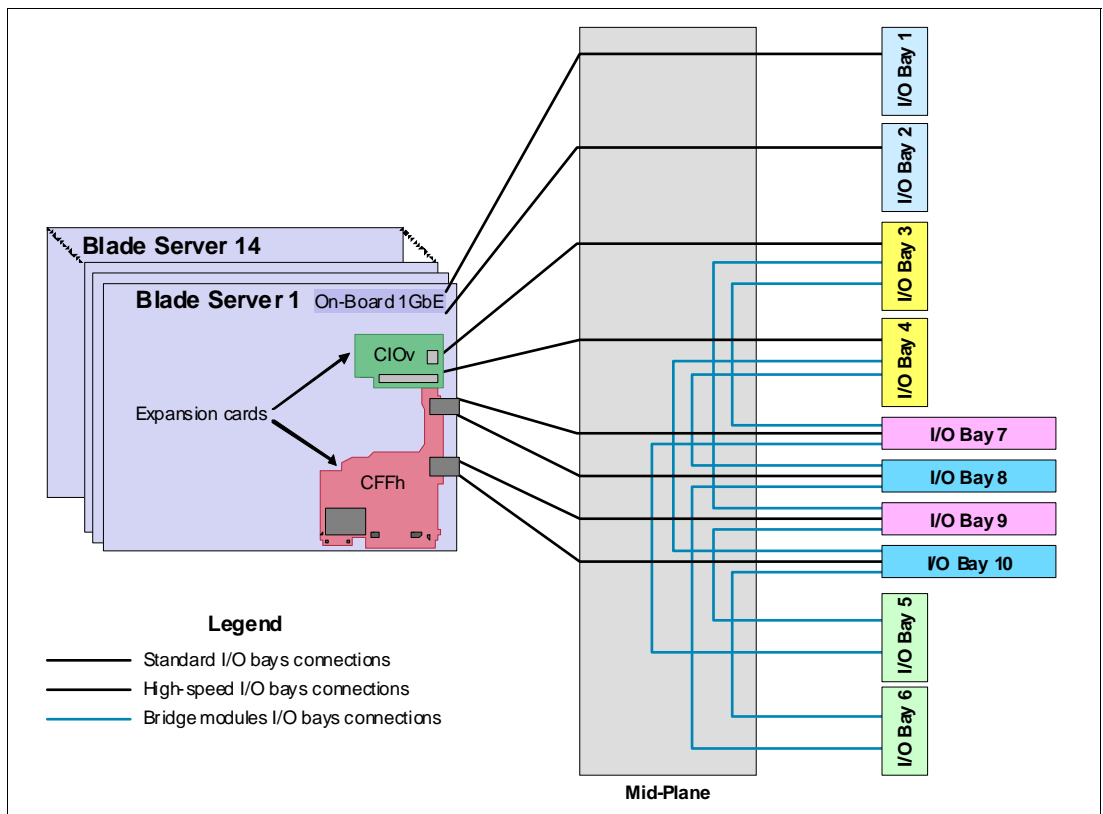


Figure 2-13 BladeCenter H I/O topology

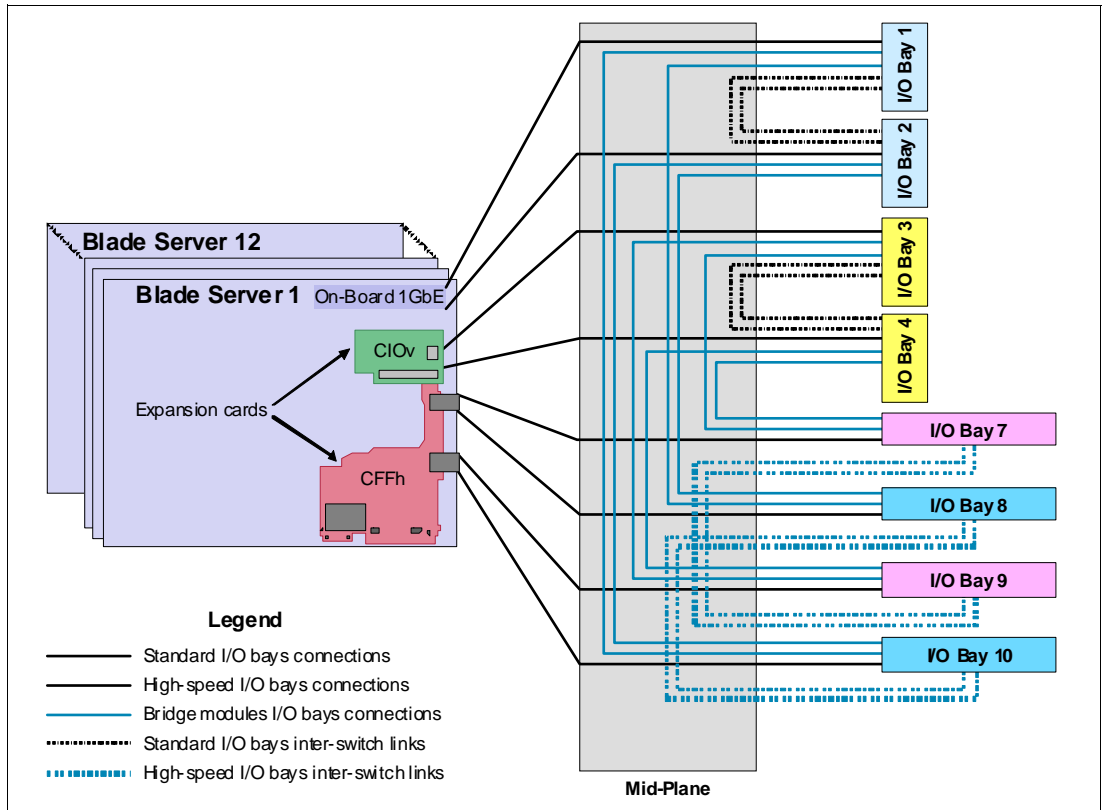


Figure 2-14 BladeCenter HT I/O topology

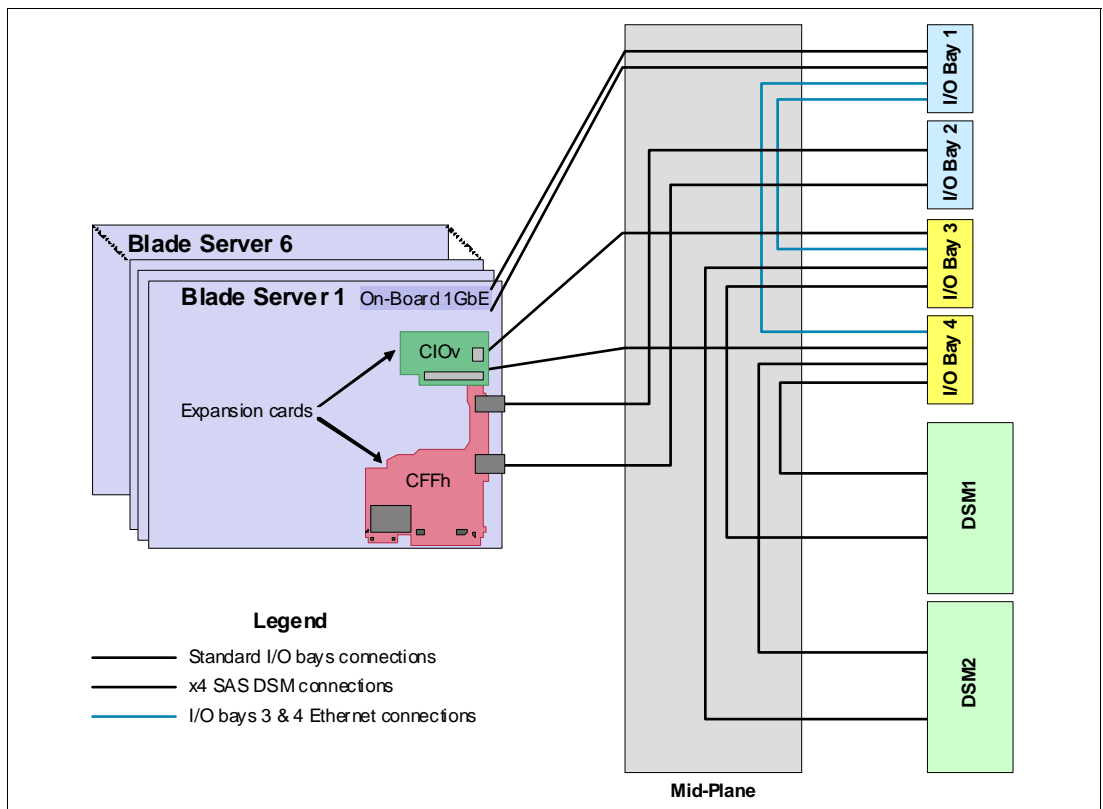


Figure 2-15 BladeCenter S I/O topology

2.6.3 I/O expansion cards

The following I/O expansion cards provide additional resources that can be used by a native operating system, the Virtual I/O Server (VIOS), or assigned directly to a LPAR by the VIOS.

See 1.5.8, “I/O features” on page 24 for details about each supported card.

LAN adapters

In addition to the onboard HEA ports, Ethernet ports can be added with LAN expansion card adapters. The current LAN adapters for the PS700, PS701, and PS702 blades are available in the CFFh form factor type. The QLogic Ethernet and 4 Gb Fibre Channel Expansion Card has two 1 Gb Ethernet ports and two 4 Gb Fibre Channel ports.

The Ethernet ports on CFFh expansion cards (BladeCenter H and HT) are connected to switch bays 7, and 9 and the Fibre Channel ports to switch bays 8, and 10. In the BladeCenter S only the Ethernet ports are usable and the connection is to Bay 2.

SAS adapter

To connect to external SAS devices, including the BladeCenter S storage modules, the 3 Gb SAS Passthrough Expansion Card and BladeCenter SAS Connectivity Modules are required.

The 3 Gb SAS Passthrough Expansion Card is a two port PICe CIOv form factor card. The output from the ports on this card are routed through the BladeCenter mid-plane to I/O switch bays 3 and 4.

Fibre Channel adapters

The PS700, PS701, and PS702 support direct or SAN connection to devices using Fibre Channel adapters and the appropriate pass-through or Fibre Channel switch modules in the BladeCenter chassis. Fibre Channel expansion cards are available in both form factors and in 4 Gb and 8 Gb data rates.

The two ports on CIOv form factor expansion cards are connected to BladeCenter I/O switch module bays 3 and 4. The two Fibre Channel ports on a CFFh expansion card connect to BladeCenter H or HT I/O switch bays 8 and 10. The Fibre Channel ports on a CFFh form factor adapters are not supported for use in a BladeCenter S chassis.

Fibre Channel over Ethernet (FCoE)

A new emerging protocol emerging, Fibre Channel over Ethernet (FCoE), is being developed within T11 as part of the Fibre Channel Backbone 5 (FC-BB-5) project. It is not meant to displace or replace FC. FCoE is an enhancement that expands FC into the Ethernet by combining two leading-edge technologies (FC and the Ethernet). This evolution with FCoE makes network consolidation a reality by the combination of Fibre Channel and Ethernet. This network consolidation maintains the resiliency, efficiency, and seamlessness of the existing FC-based data center.

Figure 2-16 on page 59 shows a configuration using BladeCenter FCoE components.

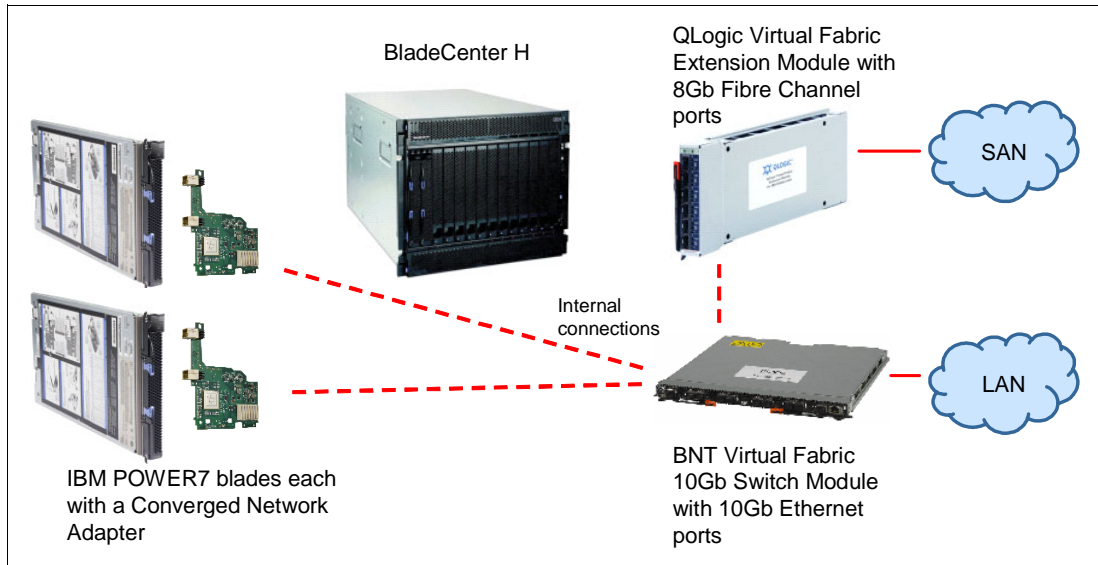


Figure 2-16 FCoE connections in IBM BladeCenter

For more information about FCoE, read *An Introduction to Fibre Channel over Ethernet, and Fibre Channel over Convergence Enhanced Ethernet*, REDP-4493, available from the following Web page:

<http://www.redbooks.ibm.com/abstracts/redp4493.html>

The QLogic 2-port 10 Gb Converged Network Adapter is a CFFh form factor card. The ports on this card are connected to BladeCenter H and HT I/O switch module bays 7 and 9. In these bays a pass through or FCoE capable I/O module can provide connectivity to a top-of-rack switch. A combination of the appropriate I/O switch module in these bays and the proper Fibre Channel capable modules in bays 3 and 5 can eliminate the top-of-rack switch requirement. See 1.6, “Supported BladeCenter I/O modules” on page 29.

InfiniBand Host Channel adapter

The InfiniBand Architecture (IBA) is an industry-standard architecture for server I/O and interserver 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 is an open set of interconnected standards and specifications. The main InfiniBand specification has been published by the InfiniBand Trade Association and is available at the following Web page:

<http://www.infinibandta.org/>

InfiniBand is based on a switched fabric architecture of serial point-to-point links. These InfiniBand 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 of 2.5–120 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, available from the following Web page:

<http://www.redbooks.ibm.com/abstracts/sg247767.html>

The 4X InfiniBand DDR Expansion Card is a 2 port CFFh form factor card and is only supported in a BladeCenter H chassis. The two ports are connected to the BladeCenter H I/O switch bays 7 and 8, and 9 and 10. These switch bays require a supported InfiniBand switch module to provide either external or blade to blade within the same chassis communication.

2.6.4 Embedded SAS Controller

The embedded SAS controller is connected to one of the 64-bit PCI-X 2.0 buses on the GX+ multifunctional host bridge chip. The PS702 uses a single embedded SAS controller. More information about the SAS I/O subsystem can be found in 2.9, “Internal storage” on page 65.

2.6.5 HEA ports

Each HEA port has its own connection to the GX+ multifunctional host bridge chip. The connections are configured for 1 Gb operation. The HEA ports are part of the Integrated Virtual Ethernet (IVE) subsystem. The IVE subsystem is described in 2.7, “Integrated Virtual Ethernet” on page 61.

2.6.6 Embedded USB controller

The USB controller is connected to the 64-bit PCI-X 1.0 bus of the GX+ multifunctional host bridge chip. This embedded USB controller provides support for four USB root ports. These ports are connected to four USB busses on the BladeCenter midplane through two connectors. Two of these ports are routed to the two AMM bays. The other two USB ports are directed to the BladeCenter Media Tray, as shown in Figure 2-17 on page 61. The PS702 uses a single embedded USB controller.

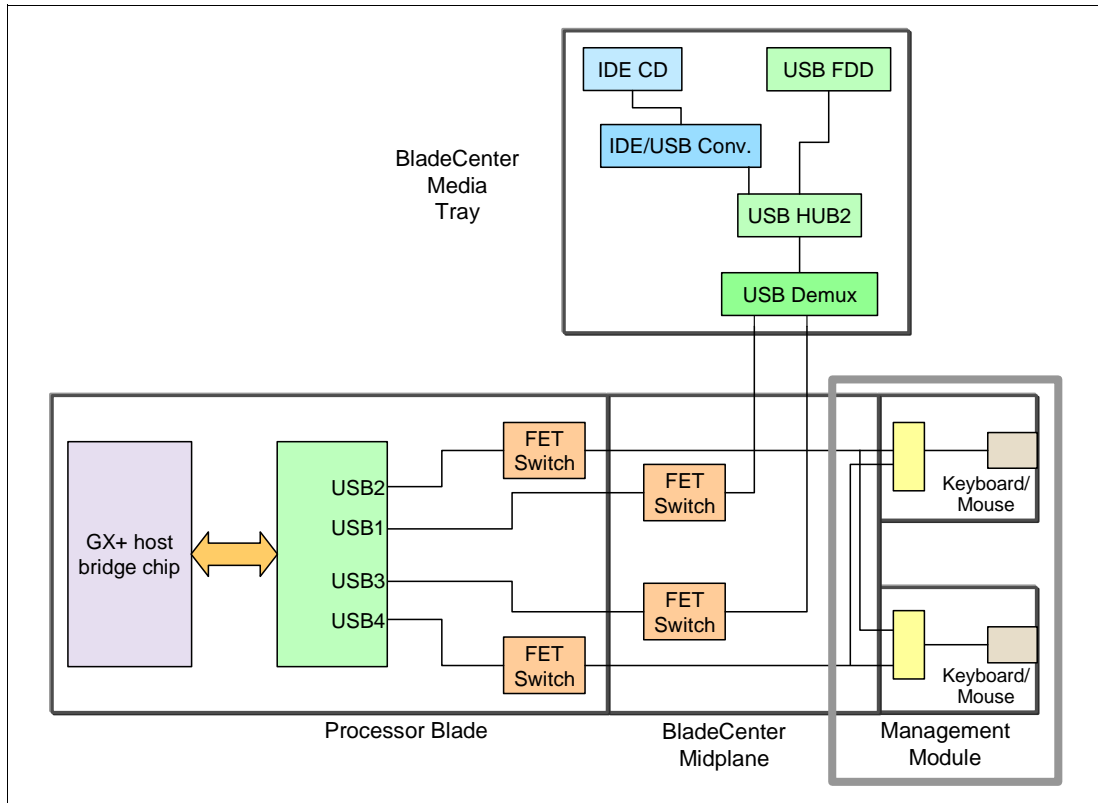


Figure 2-17 System overview of USB connections

Note: The PS700, PS701, and PS702 blades do not support the KVM function from the AMM.

BladeCenter Media Tray

The BladeCenter Media Tray, depending on the BladeCenter chassis used, can contain up to two USB ports, one optical drive and system status LEDs. For information about the different media tray options available by BladeCenter model see *IBM BladeCenter Products and Technology*, SG24-7523 available from the following Web page:

<http://www.redbooks.ibm.com/abstracts/sg247523.html>

The media tray is a shared resources that can be assigned to any blade slot.

2.7 Integrated Virtual Ethernet

Introduced with POWER6, POWER7 processor-based servers continue the use of IVE. The terms IVE and HEA are sometimes used interchangeably, however, IVE encompasses all the hardware parts including the HEA and the integration of several technologies.

IVE enables the ability to manage the sharing of the integrated HEA physical ports. The PS700, PS701, and PS702 blades include two 1 Gb HEA ports on the base blade. The PS702 has an additional two HEA ports on the expansion unit for a total of four physical ports.

IVE provides logical Ethernet ports that can communicate to logical partitions (LPARs) reducing the use of IBM POWER Hypervisor™. The design provides a logical connection for multiple LPARs to a physical port, allowing LPARs to access external networks through the HEA without using a Shared Ethernet Adapter (Ethernet bridge) through the Virtual I/O Server. 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 HEA ports with improved performance.

Figure 2-18 shows the difference between IVE and SEA implementations.

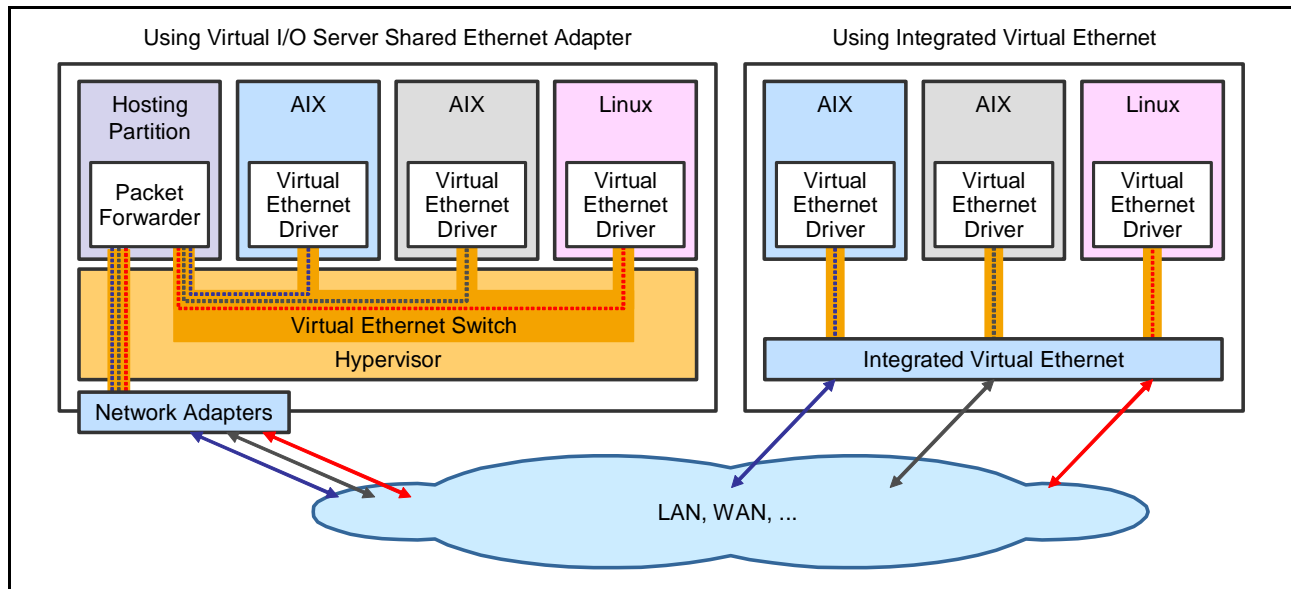


Figure 2-18 IVE compared to Virtual I/O Server Shared Ethernet Adapter

IVE design meets general market requirements for better performance and better virtualization for Ethernet. It offers the following benefits:

- ▶ Either two 1 Gbps HEA ports (PS700 and PS701) or four 1 Gbps HEA ports (PS702)
- ▶ Logical ports assigned to LPARs for external network connectivity as an option to a Virtual I/O Server provided Shared Ethernet Adapter (SEA)
- ▶ Industry standard hardware acceleration, loaded 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

For more information about IVE features read *Integrated Virtual Ethernet Adapter Technical Overview and Introduction*, REDP-4340, available at the following Web page:

<http://www.redbooks.ibm.com/abstracts/redp4340.html>

2.7.1 IVE subsystem

One of the key design goals of the IVE architecture 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. Any 10 Gbps, 1 Gbps, 100 Mbps, or 10 Mbps speeds share the same I/O.

The IVE implementation on the PS700, PS701, and PS702 blades use a maximum rate of 1 Gbps and the HEA ports are integrated onto the base blades and expansion unit. The two physical ports on the PS700 and PS701 are associated to a single logical port group. The two additional physical ports on the PS702 expansion unit are associated to a second port group. Each port group can address up to 16 logical ports, A maximum of 16 MAC addresses are assigned to a port group. A maximum of one logical port per physical port can be given to a LPAR.

The HEA ports are connected to the BladeCenter I/O switch module bays 1 and 2, with the exception of the BladeCenter S. The BladeCenter S connects all HEA ports to I/O switch module bay 1.

Note: On PS700, PS701, and PS702 blades, as of this writing, Virtual I/O Server must be installed on the blade to configure IVE logical ports through Integrated Virtualization Manager (IVM). Native operating system installations can only use the physical ports.

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. Flash code update, therefore, is done by the POWER Hypervisor.

Important: The HEA port implementation on the PS700, PS701, and PS702 blades always shows a link status of up, and should be considered when implementing network failover scenarios.

2.8 Service processor

The service processor (previously known as the Flexible Service Processor or FSP) is used to monitor and manage the system hardware resources and devices. In a POWER7-based blade implementation the external network connection for the service processor is routed through an on-blade Ethernet switch, through the BladeCenter midplane, chassis switches and to the AMM. The Serial over LAN (SOL) connection for a system console uses this same connection. When the blade is in standby power mode the service processor responds to AMM instructions and can detect Wake-on-LAN (WOL) packets.

The PS700 and PS701 each have a single service processor. The PS702 has a second service processor in the expansion unit. However, it is only used for controlling and managing the hardware on this second planar.

2.8.1 Server console access by SOL

The PS700, PS701, and PS702 blades do not have an on-board video chip and do not support KVM connections. Server console access is obtain by a SOL connection.

SOL provides a means to manage servers remotely by using a command-line interface (CLI) over a Telnet or secure shell (SSH) connection. SOL is required to manage servers that do not have KVM support. SOL provides console redirection for both System Management Services (SMS) and the blade server operating system. The SOL feature redirects server serial-connection data over a LAN without requiring special cabling. The SOL connection enables blade servers to be managed from any remote location with network access.

SOL offers the following advantages:

- ▶ Remote administration without keyboard, video, or mouse (headless servers)
- ▶ Reduced cabling and without requiring a serial concentrator
- ▶ Standard Telnet interface, eliminating the requirement for special client software

The IBM BladeCenter AMM CLI provides access to the text-console command prompt on each blade server through a SOL connection, enabling the blade servers to be managed from a remote location.

In the BladeCenter environment, the SOL console data stream from a blade is routed from the blades's service processor to the AMM through the on-blade switch to the network infrastructure of the BladeCenter unit, including an Ethernet-compatible I/O module that supports SOL communication. Figure 2-19 shows the SOL data stream flow.

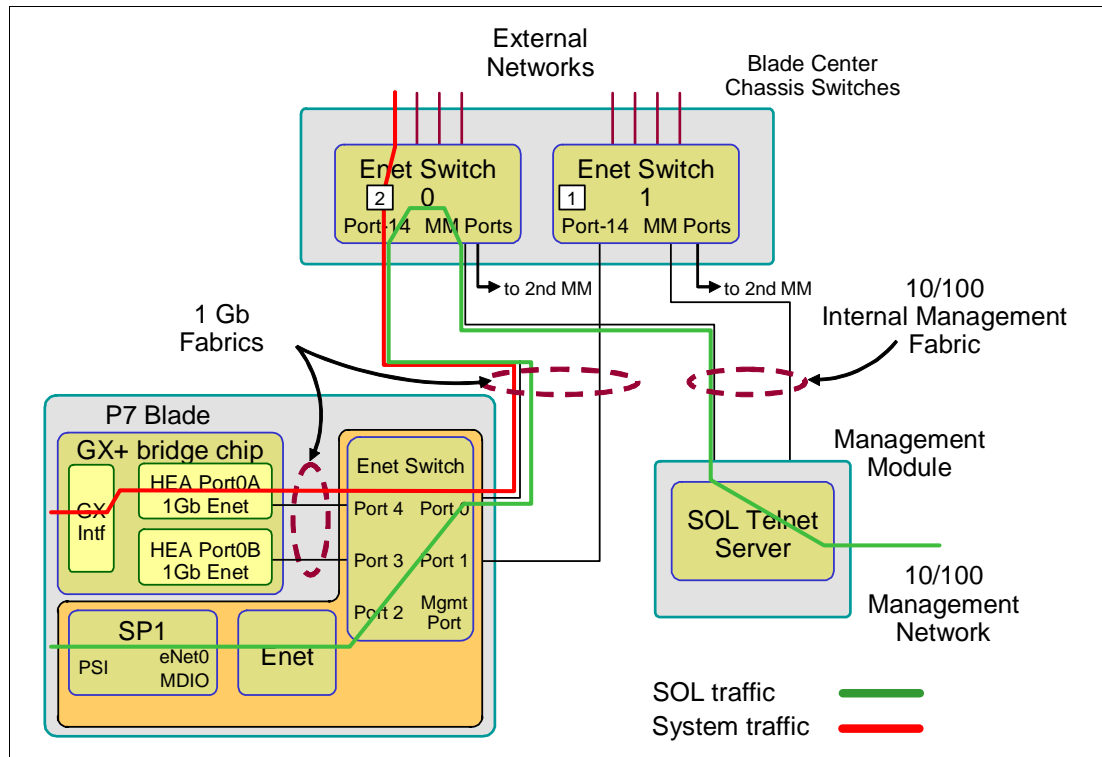


Figure 2-19 SOL service processor to AMM connection

BladeCenter components are configured for SOL operation through the BladeCenter AMM. The AMM also acts as a proxy in the network infrastructure to couple a client running a Telnet or SSH session with the management module to an SOL session running on a blade server, enabling the Telnet or SSH client to interact with the serial port of the blade server over the network.

Because all SOL traffic is controlled by and routed through the AMM, administrators can segregate the management traffic for the BladeCenter unit from the data traffic of the blade servers. To start an SOL connection with a blade server, perform the following steps:

1. Start a Telnet or SSH CLI session with the AMM.
2. Start a remote-console SOL session with any blade server in the BladeCenter unit that is set up and enabled for SOL operation.

You can establish up to 20 separate Web-interface, Telnet, or SSH sessions with a BladeCenter AMM. For a BladeCenter unit, this step enables you to have 14 simultaneous SOL sessions active (one for each of up to 14 blade servers) with six additional CLI sessions available for BladeCenter unit management.

With a BladeCenter S unit you have six simultaneous SOL sessions active (one for each of up to six blade servers) with 14 additional CLI sessions available for BladeCenter unit management. If security is a concern, you can use Secure Shell (SSH) sessions, or connections made through the serial management port that is available on the AMM, to establish secure Telnet CLI sessions with the BladeCenter management module before starting an SOL console-redirect session with a blade server.

SOL has the following requirements:

- ▶ An Ethernet switch module or Intelligent Pass-Thru Module is installed in bay 1 of a BladeCenter
- ▶ SOL is enabled for those blades that you want to connect to with SOL.
- ▶ The Ethernet switch module must be set up correctly.

For details about setting up SOL, see the *BladeCenter Serial Over LAN Setup Guide*, which can be found at the following Web page:

<http://www.ibm.com/support/docview.wss?uid=psg1MIGR-54666>

This guide contains an example of how to establish a Telnet or SSH connection to the management module and then an SOL console.

2.9 Internal storage

PS700, PS701 and PS702 blades use an integrated SAS controller. The controller's PCI-X interface to the GX+ multifunctional host bridge chip is 64 bits wide and operates at 133 MHz.

This controller provides ports for the internal drives, and ports through the 3 Gb SAS Passthrough Expansion Card to the BladeCenter SAS switch modules. The SAS controller ports used for the internal disk drives can support a single 2.5" SAS hard disk drive (HDD) at each DASD bay location, as shown in Figure 2-20 on page 66, Figure 2-21 on page 66. and Figure 2-22 on page 67.

Note: Solid state drives (SSDs) are not supported.

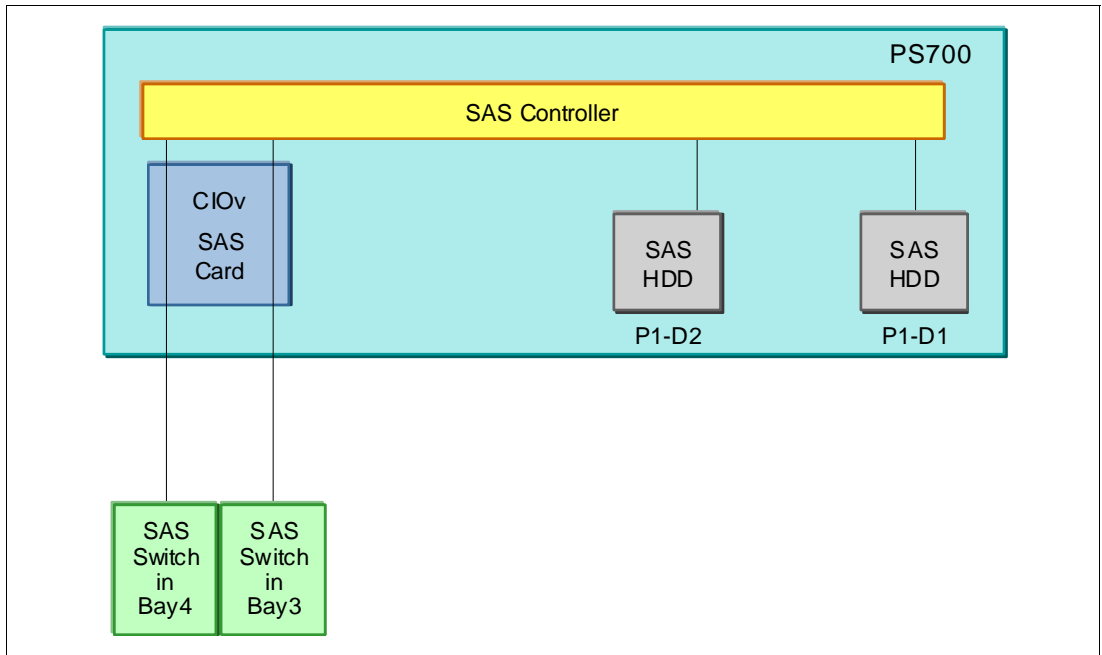


Figure 2-20 PS700 SAS configuration

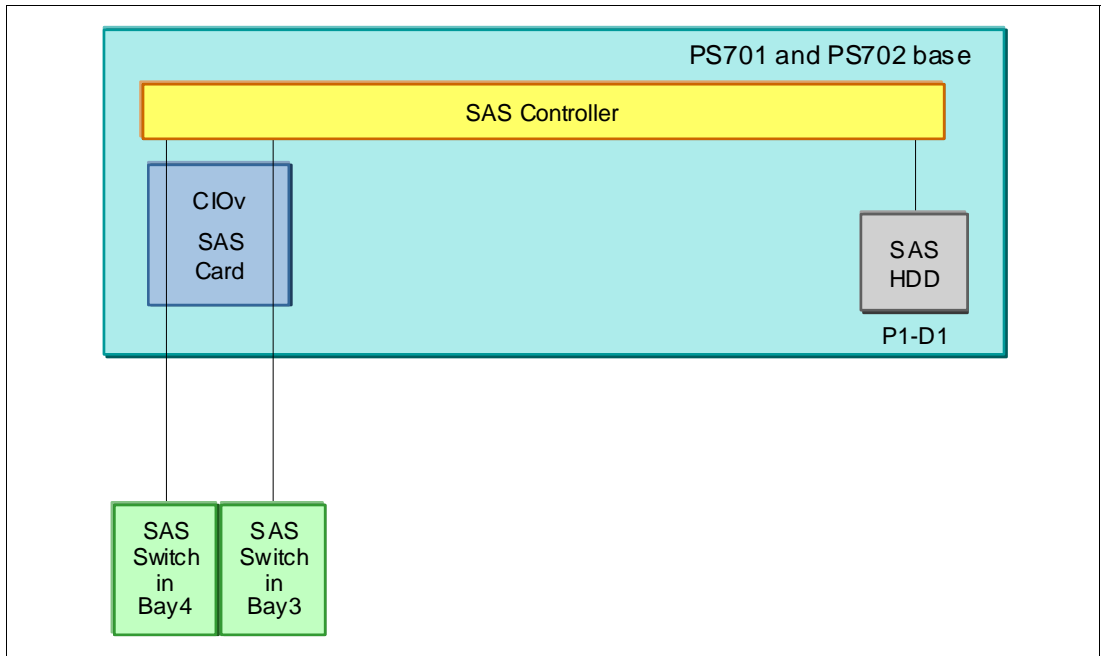


Figure 2-21 PS701 SAS configuration

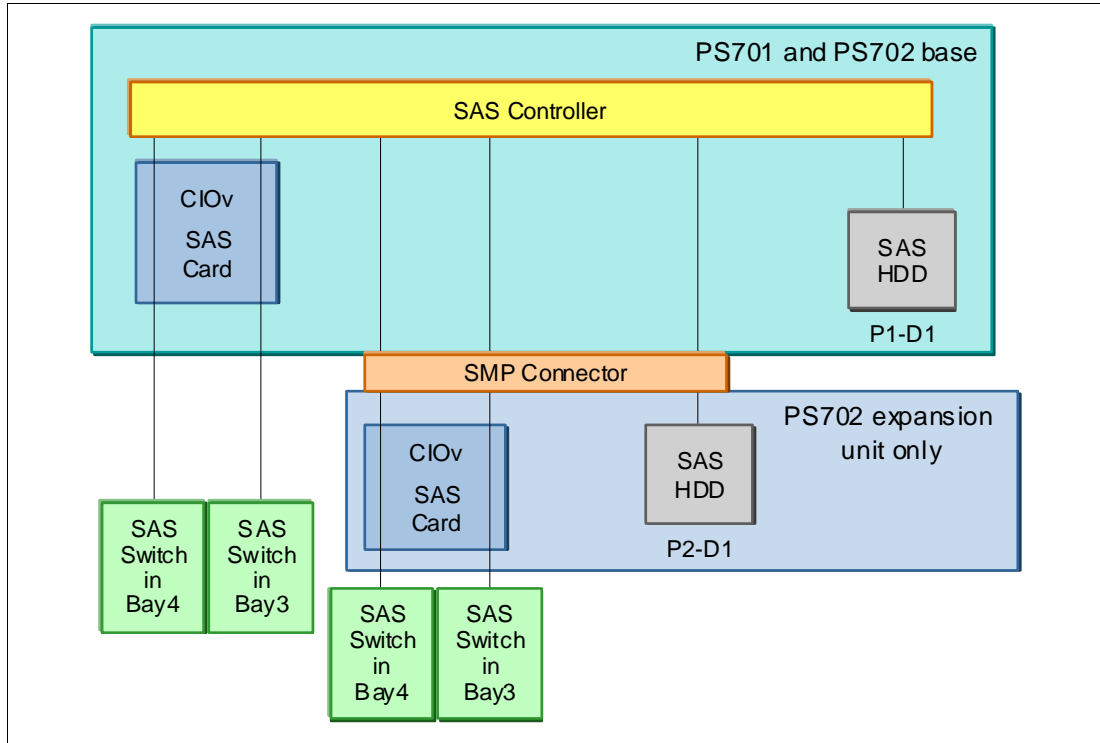


Figure 2-22 PS702 SAS configuration

Figure 2-23 show the physical locations and codes for the HDDs in the PS700

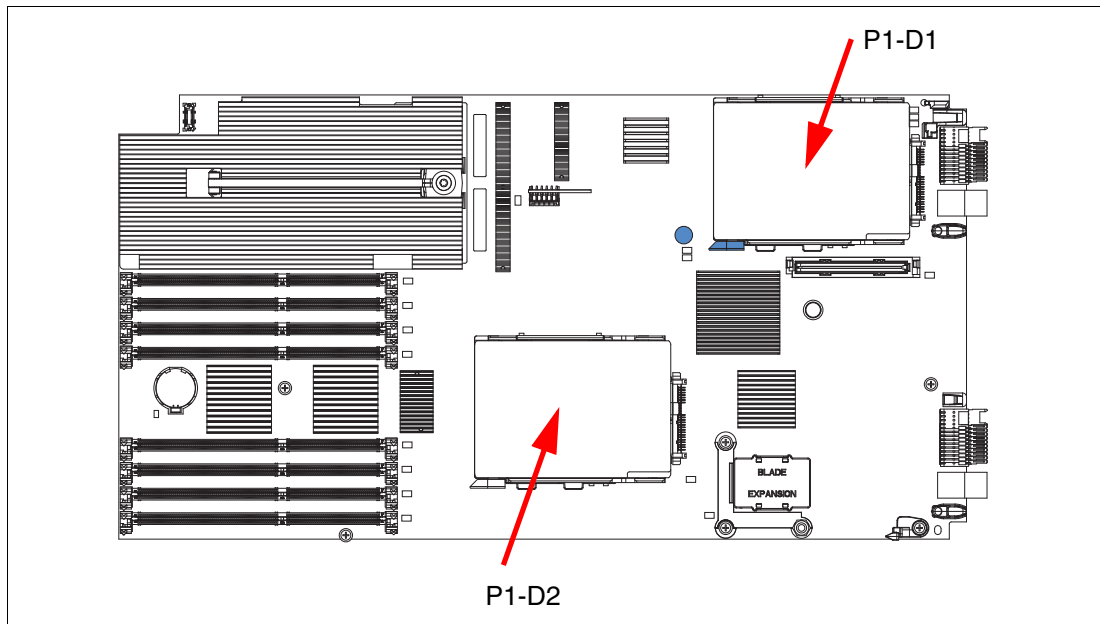


Figure 2-23 HDD location and physical location code PS700

Figure 2-24 show the physical location and code for a HDD in a PS701. The PS702 expansion unit locates the HDD is the same location with a physical location code of P2-D1.



Figure 2-24 HDD location and physical location code PS701

2.9.1 Hardware RAID function

The PS700 and PS702 have support for RAID functions across a blade's internal when more than one storage drive is installed in the system through the SAS controller. RAID 0 and RAID 1 are supported. If there is only one drive, there is no RAID function. The PS701 only supports one drive so RAID is not offered.

The configuration of the RAID array the blade's internal disks is performed by booting the system from the AIX Diagnostic Utilities disk prior to installing the operating system.

2.9.2 External SAS connections

The onboard SAS controller in the PS700, PS701, and PS702 blades does not provide a direct access external SAS port. However, by using a 3 Gb SAS Passthrough Expansion Card and BladeCenter SAS Connectivity Modules, two ports on the SAS controller (four in the PS702 with a second SAS card on the expansion unit) are expanded, providing access to BladeCenter S Disk Storage Modules (DSM) or an external SAS disk sub-system.

2.10 External disk subsystems

This section describes the external disk subsystems, supported IBM System Storage family of products.

For up-to-date compatibility information for Power blades and IBM Storage, go to the the Storage System Interoperability Center at the following link:

<http://ibm.com/systems/support/storage/config/ssic>

For N Series Storage compatibility with Power blades, go to:

<http://ibm.com/systems/storage/network/interophome.html>

2.10.1 IBM BladeCenter S Disk Storage Modules

The BladeCenter S supports up to two storage modules. These modules provide integrated SAS storage functionality to the BladeCenter S chassis.

The storage module's collection of disk drives are made accessible to blade servers through a SAS Connectivity Module or SAS RAID Controller Module installed in the BladeCenter S chassis and SAS expansion cards installed in the blades. The SAS RAID Controller Module provides RAID 0, 1, 5, and 10 support.

Each of the two storage modules contain up to six 3.5 inch hot-swap hard drives, for a total of 12 internal drives. The storage module supports SAS, SATA, and Near Line SAS (NL SAS) drives. Intermixing SAS and SATA or SAS and NL SAS drives within the same storage module is supported.

2.10.2 IBM System Storage

The IBM System Storage Disk Systems products and offerings provide compelling storage solutions with superior value for all levels of business.

IBM System Storage N series

IBM N series unified system storage solutions can provide customers with the latest technology to help them improve performance, virtualization manageability, and system efficiency at a reduced total cost of ownership. Several enhancements have been incorporated to the N series product line, to complement and reinvigorate this portfolio of solutions:

- ▶ The new SnapManager® for Hyper-V provides extensive management for backup, restoration, and replication for Microsoft® Hyper-V environments
- ▶ The new N series Software Packs provides the benefits of a broad set of N series solutions at a reduced cost.
- ▶ An essential component to this launch is Fibre Channel over Ethernet access and 10 Gb Ethernet, to help integrate Fibre Channel and Ethernet flow into a unified network, and take advantage of current Fibre Channel installations.

For more information, see the following Web page:

<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 FlashCopy® images per volume provide a reliable virtualization platform with the support of Microsoft Windows® Server 2008 with HyperV.

For more information, see the following Web page:

<http://www.ibm.com/systems/storage/disk/ds3000/>

IBM System Storage DS5020 Express

Optimized data management requires storage solutions with high data availability, strong storage management capabilities and powerful performance features. IBM offers the IBM System Storage DS5020 Express, designed to provide lower total cost of ownership, high

performance, robust functionality, and unparalleled ease of use. As part of the IBM DS series, the DS5020 Express offers the following features:

- ▶ High-performance 8 Gbps capable Fibre Channel connections
- ▶ Optional 1 Gbps iSCSI interface
- ▶ Up to 112 TB of physical storage capacity with 112 1 TB SATA disk drives
- ▶ Powerful system management, data management, and data protection features

For more information, see the following Web page:

<http://www.ibm.com/systems/storage/disk/ds5020/>

IBM System Storage DS5000

New DS5000 enhancements help reduce costs by reducing power per performance by introducing SSD drives. Also, with the new EXP5060 expansion unit supporting 60 1-TB SATA drives in a 4 U package, you can see up to a one-third reduction in floor space over standard enclosures. With the addition of 1 Gbps iSCSI host-attach, you can reduce cost for less demanding applications and continue providing high performance where necessary by using the 8 Gbps FC host ports. With DS5000, you get consistent performance from a smarter design, that simplifies your infrastructure, improves your total cost of ownership (TCO), and reduces costs.

For more information, see the following Web page:

<http://www.ibm.com/systems/storage/disk/ds5000>

IBM XIV Storage System

IBM is introducing a mid-sized configuration of its self-optimizing, self-healing, resilient disk solution, the IBM XIV® Storage System. Organizations with mid-sized capacity requirements can take advantage of the latest technology from IBM for their most demanding applications with as little as 27 TB of usable capacity and incremental upgrades.

For more information, see the following Web page:

<http://www.ibm.com/systems/storage/disk/xiv/>

IBM System Storage DS8700

The IBM System Storage DS8700 is the most advanced model in the IBM DS8000 lineup and introduces dual IBM POWER6 based controllers that usher in a new level of performance for the company's flagship enterprise disk platform. The new DS8700 supports the most demanding business applications with its superior data throughput, unparalleled resiliency features and five-nines availability. In today's dynamic, global business environment, where organizations such as yours need information be reliably available around the clock and with minimal delay, can you really afford not to run your business on the DS8000 series? With its tremendous scalability, flexible tiered storage options, broad server support, and support for advanced IBM duplication technology, the DS8000 can help simplify the storage environment by consolidating multiple storage systems onto a single system, and provide the availability and performance you have come to trust for your most important business applications.

For more information, see the following Web page:

<http://www.ibm.com/systems/storage/disk/ds8000/>

2.11 IVM

IVM is a simplified hardware management solution that is part of the PowerVM implementation on the PS700, PS701, and PS702 blades. POWER processor-based blades do not include an option for attachment to a Hardware Management Console (HMC).

IVM inherits most of the HMC features and capabilities and enables the exploitation of PowerVM technology. It manages a single server, avoiding the need for 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 is an addition to the Virtual I/O Server, 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. See Table 2-9.

For a complete description of the possibilities offered by IVM, see *Integrated Virtualization Manager on IBM System p5*, REDP-4061, available the following Web page:

<http://www.redbooks.ibm.com/abstracts/redp4061.html>

Table 2-9 Comparison of IVM and HMC

| Characteristic | IVM | HMC |
|-----------------------------------|---|---|
| General characteristics | | |
| Delivery vehicle | Integrated into the server | A desktop or rack-mounted appliance |
| Footprint | Runs in 60 MB memory and requires minimal CPU as it runs stateless. | 2-Core x86, 2 GB RAM, 80 GB HD |
| Installation | Installed with the Virtual I/O Server (optical or network). Preinstall option available on certain systems. | Appliance is preinstalled. Reinstall through optical media or network is supported. |
| Multiple system support | One IVM per server | One HMC can manage multiple servers (48 CECs / 1024 LPARS) |
| User interface | Web browser (no local graphical display) and telnet session | Web browser (local or remote) |
| Scripting and automation | VIOS command-line interface (CLI) and HMC compatible CLI. | HMC CLI |
| RAS characteristics | | |
| Redundancy and HA of manager | Only one IVM per server | Multiple HMCs can manage the same system for HMC redundancy. |
| Multiple VIOS | No, single VIOS | Yes |
| Fix or update process for manager | VIOS fixes and updates | HMC e-fixes and release updates |
| Adapter microcode updates | Inventory scout through RMC | Inventory scout through RMC |
| Firmware updates | Inband through OS; not concurrent | Service Focal Point™ with concurrent firmware updates |

| Characteristic | IVM | HMC |
|--|--|---|
| I/O concurrent maintenance (not available on POWER based blades) | VIOS support for slot and device level concurrent maintenance through 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 |
| Support for multiple profiles per partition | No | Yes |
| SysPlan Deploy & mksysplan | No | No |

2.12 Operating system support

The IBM POWER7 processor-based systems supports 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, or PowerVM Active Memory™ Sharing.

Note: For details about the software available on IBM POWER servers, see Power Systems Software™ at the following Web page:

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

The PS700, PS701, and PS702 blades support the following operating system versions.

Virtual I/O Server

Virtual I/O Server 2.1.3.0 or later

IBM regularly updates the Virtual I/O Server code. To find information about the latest updates, see the Virtual I/O Server at the following Web page:

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

IBM AIX Version 5.3

AIX Version 5.3 with the 5300-12 Technology Level or later.

A partition using AIX Version 5.3 executes in POWER6 or POWER6+ compatibility mode.

IBM periodically releases maintenance packages (service packs or technology levels) for the AIX 5L operating system. Information about these packages, downloading, and obtaining the CD-ROM is on the Fix Central Web page:

<http://www.ibm.com/eserver/support/fixes/fixcentral/main/pseries/aix>

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

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

AIX Version 6.1

AIX 6.1 with the 6100-05 Technology Level or later

For information regarding AIX V6.1 maintenance and support, go to the Fix Central Web page:

<http://www.ibm.com/eserver/support/fixes/fixcentral/main/pseries/aix>

IBM i

Virtual I/O Server is required to install IBM i in a LPAR on PS700, PS701, and PS702 blades and all I/O must be virtualized.

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

For a detailed guide on installing and operating IBM i with Power Blades, see the following Web page:

http://ibm.com/systems/resources/systems_power_hardware_blades_i_on_blade_readme.pdf

Linux

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

At the time of this writing, the supported versions of Linux on POWER7 processor technology based servers are as follows:

- ▶ SUSE Linux Enterprise Server 10 with SP3 and the latest maintenance, in POWER6 Compatibility mode
- ▶ SUSE Linux Enterprise Server 11 with SP1 or later, supporting POWER6 or POWER7 mode
- ▶ Red Hat RHEL 5.5 in POWER6 Compatibility mode

Linux operating system licenses are ordered separately from the hardware. You can obtain Linux operating system licenses from IBM, to be included with your POWER7 processor technology-based servers, or from other Linux distributors.

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

<http://www.ibm.com/systems/p/os/linux/>

For information about SUSE Linux Enterprise Server, go to the following Web page:

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

For information about Red Hat Enterprise Linux Advanced Server, go to the following Web page:

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

Supported virtualization features are listed in 3.3.8, “Supported PowerVM features by operating system” on page 98.

2.13 IBM EnergyScale

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

The BladeCenter AMM and IBM Systems Director Active Energy Manager exploit EnergyScale technology, enabling advanced energy management features to conserve power and 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 reduce frequency to save energy.

2.13.1 IBM EnergyScale technology

This section describes IBM EnergyScale design features, and hardware and software requirements.

IBM EnergyScale consists of the following elements:

- ▶ A built-in EnergyScale device (formally known as Thermal Power Management Device or TPMD)
- ▶ Power executive software. IBM Systems Director Active Energy Manager, an IBM Systems Directors plug-in and BladeCenter AMM.

IBM EnergyScale functions include the following elements:

- ▶ Energy trending

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

- ▶ Thermal reporting

IBM Systems 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 require attention.

- ▶ Power Saver Mode

Power Saver Mode reduces the processor frequency and voltage by a fixed amount, reducing the energy consumption of the system and 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 50% from nominal. Power Saver Mode is not supported during boot or reboot operations, although it is a persistent condition that is sustained after the boot when the system starts executing instructions.

- ▶ Dynamic Power Saver Mode

Dynamic Power Saver Mode varies processor frequency and voltage based on the use of the POWER7 processors. The user must configure this setting from the BladeCenter AMM or IBM Director Active Energy Manager. Processor frequency and use are inversely proportional for most workloads, implying that as the frequency of a processor increases, its use 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 use. When a system is idle, the system firmware lowers the frequency and voltage to Power Saver Mode values. When fully used, the maximum frequency varies, depending on whether the user favors power savings or system performance. If an administrator prefers energy savings and a system is fully-used, the system can reduce the maximum frequency to 95% of nominal values. If performance is favored over energy consumption, the maximum frequency will be at least 100% of nominal. 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 throttling the processors in the system, degrading performance significantly. The idea of a power cap is to set a limit that should never be reached but frees up margined power in the data center. The margined power is the amount of extra power that is allocated to a server during its installation in a datacenter. It is based on the server environmental specifications that usually are never reached because server specifications are always based on maximum configurations and worst case scenarios. The user must set and enable an energy cap from the BladeCenter AMM or IBM Systems Director Active Energy Manager user interface.

- ▶ 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 an energy management goal is to meet a particular consumption limit, soft power capping is the mechanism to use.

- ▶ Processor Core Nap

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 falls within a partition dispatch (context switch) such that the POWER Hypervisor can use it 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 clock-off 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. Unlicensed cores are kept in core Nap until they are licensed and return to core Nap when they are unlicensed again.

- ▶ 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 longer.

- ▶ EnergyScale for I/O

IBM POWER processor-based systems automatically power off 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 POWER processor-based servers, and the expansion units that they support.

In addition to the normal EnergyScale functions, the EnergyScale device in the PS700, PS701, and PS702 blades incorporate the following BladeCenter functions:

- ▶ Transition from over-subscribed power consumption to nominal power consumption when commanded by the BladeCenter AMM. This transition is signaled by the AMM as a result of a redundant power supply failure in the BladeCenter.
- ▶ Report blade power consumption to the AMM through the service processor
- ▶ Report blade system voltage levels to the AMM through the service processor
- ▶ Accommodate BladeCenter/AMM defined thermal triggers such as warning temperature, throttle temperature, and critical temperature

2.13.2 EnergyScale device

The EnergyScale device dynamically optimizes the processor performance depending on processor power and system workload.

The IBM POWER7 chip is a significant improvement in power and performance over the IBM POWER6 chip. POWER7 has more internal hardware, and power and thermal management functions to interact with:

- ▶ More hardware: Eight cores versus two cores, four threads versus two threads per core, and asynchronous processor core chiplet
- ▶ Advanced Idle Power Management functions
- ▶ Advanced Dynamic Power Management (DPM) functions in all units in hardware (processor cores, processor core chiplet, chip-level nest unit level, and chip level)
- ▶ Advanced Actuators/Control
- ▶ Advanced Accelerators

The new EnergyScale device has a more powerful microcontroller, more A/D channels and more busses to handle the increase workload, link traffic, and new power and thermal functions.



Virtualization

IBM Advance POWER Virtualization (PowerVM) is a feature used to consolidate workload to deliver cost savings and improve infrastructure responsiveness. As we look for ways to maximize the return on your IT infrastructure investments, consolidating workloads and increasing server use becomes an attractive proposition.

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

- ▶ Improve server use by consolidating diverse sets of applications.
- ▶ Share CPU, memory, and I/O resources to reduce total cost of ownership.
- ▶ Improve business responsiveness and operational speed by dynamically re-allocating resources to applications as needed, to better anticipate changing business needs.
- ▶ Simplify IT infrastructure management by making workloads independent of hardware resources, enabling you to make business-driven policies to deliver resources based on time, cost, and service-level requirements.
- ▶ Move running workloads between servers to maximize availability and avoid planned downtime

This chapter discusses the virtualization technologies and features on IBM POWER7 processor-based blade servers:

- ▶ 3.1, “POWER Hypervisor” on page 78
- ▶ 3.2, “POWER processor modes” on page 82
- ▶ 3.3, “PowerVM” on page 83

3.1 POWER Hypervisor

Combined with features designed into the POWER7 processors, the POWER Hypervisor delivers functions that enable capabilities, including dedicated processor partitioning, micro-partitioning, virtual processors, IEEE VLAN compatible virtual switch, and virtual SCSI adapters, virtual Fibre Channel adapters, and virtual consoles.

The POWER Hypervisor technology is integrated with all IBM POWER servers including the POWER7 processor-based blade servers. The hypervisor orchestrates and manages system virtualization, including creating logical partitions and dynamically moving resources across multiple operating environments. The POWER Hypervisor is a basic component of the system firmware that is layered between the hardware and operating system. POWER Hypervisor offers the following functions:

- ▶ Provides an abstraction layer between the physical hardware resources and the logical partitions using them
- ▶ Enforces partition integrity by providing a security layer between logical partitions
- ▶ Controls the dispatch of virtual processors to physical processors and 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 Ethernet switch between logical partitions that help to reduce the need for physical Ethernet adapters for interpartition communication
- ▶ Monitors the service processor and performs a reset or reload if it detects the loss of the service processor, notifying the operating system if the problem is not corrected
- ▶ Uses micro-partitioning to allow multiple instances of operating system to run on POWER6 and POWER7 processor-based servers or Blades

The POWER Hypervisor is always installed and activated, regardless of system configuration. The POWER Hypervisor does not own any physical I/O devices. All physical I/O devices in the system are owned by logical partitions or the Virtual I/O Server.

Memory is required to support the resource assignment to the logical partitions on the server. The amount of memory required by the POWER Hypervisor firmware varies according to several factors. The following factors influence 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 system, as shown in Table 3-1.

Table 3-1 Configured CEC memory-to-default LMB size

| Configurable memory in the system | Default Logical Memory Block |
|--|-------------------------------------|
| Less than 4 GB | 16 MB |
| Greater than 4 GB up to 8 GB | 32 MB |
| Greater than 8 GB up to 16 GB | 64 MB |
| Greater than 16 GB up to 32 GB | 128 MB |
| Greater than 32 GB | 256 MB |

Physical memory assigned to partitions are in increments of LMB.

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

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

Virtual I/O adapters are defined by system administrators during logical partition definition. Configuration information for the adapters is presented to the partition operating system.

Virtual SCSI

The POWER Hypervisor provides a virtual SCSI mechanism for virtualization of storage devices. Virtual SCSI allows secure communications between a logical partition and the IO Server (VIOS). The storage virtualization is accomplished by pairing two adapters: a virtual SCSI server adapter on VIOS and a virtual SCSI client adapter on IBM i, Linux, or AIX partitions. The combination of Virtual SCSI and VIOS provides the opportunity to share physical disk adapters in a flexible and reliable manner.

Virtual Ethernet

The POWER Hypervisor provides an IEEE 802.1Q VLAN-style virtual Ethernet switch that allows partitions on the same server to use a fast and secure communication without any need for physical connection.

Virtual Ethernet support starts with AIX Version 5.3, or the appropriate level of Linux supporting virtual Ethernet devices (see 3.3.8, “Supported PowerVM features by operating system” on page 98). 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 (=65408 -14 for the header) in non-VLAN case and to 65390 (=65408-14- 4) 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).
- ▶ An AIX partition supports 256 virtual Ethernet adapters for each logical partition. 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 operating system partition 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 VIOS partition (see 3.3.3, “VIOS” on page 88 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 for inter-partition communication.

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 VIOS logical partition. The VIOS logical partition provides the connection between the virtual Fibre Channel adapters on the VIOS logical partition and the physical Fibre Channel adapters on the managed system.

NPIV is a standard technology for Fibre Channel networks. It enables you to connect multiple logical partitions to one physical port of a physical Fibre Channel adapter. Each logical partition is identified by a unique WWPN, which means that you can connect each logical partition to independent physical storage on a SAN.

Note: To enable NPIV on a managed system, we need VIOS to be at version 2.1 or later. Also, check if the Fibre Channel adapter on managed system supports NPIV.

You can only configure virtual Fibre Channel adapters on client logical partitions that run the following operating systems:

- ▶ AIX version 6.1 Technology Level 2, or later
- ▶ AIX 5.3 Technology Level 9
- ▶ IBM i version 6.1.1, or later
- ▶ SUSE Linux Enterprise Server 11, or later

For details on which expansion card support NPIV see 3.3.7, “N_Port ID Virtualization (NPIV)” on page 96

On systems that are managed by the Integrated Virtualization Manager (IVM), you can dynamically add and remove worldwide port names (WWPNs) to and from logical partitions, and you can dynamically change the physical ports to which the WWPNs are assigned. You can also view information about the virtual and physical Fibre Channel adapters and the WWPNs by using the **lsmap** and **lsnports** commands.

For more information about how virtual Fibre Channel is managed on IVM see the following Web page:

<http://publib.boulder.ibm.com/infocenter/powersys/v3r1m5/index.jsp?topic=/arecu/arecukickoff.htm>

Figure 3-1 on page 81 depicts the connections between the client partition virtual Fibre Channel adapters and the external storage.

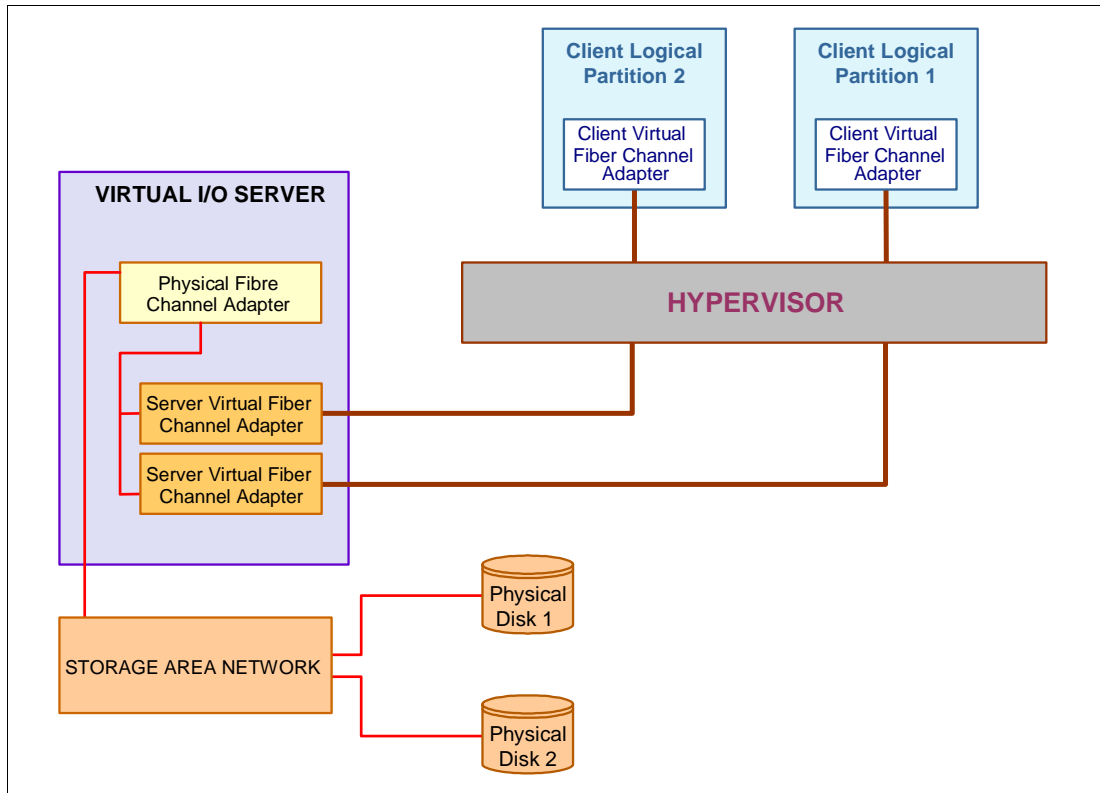


Figure 3-1 Connectivity between virtual Fibre Channels adapters and external SAN devices

Virtual Serial Adapters (TTY) console

Virtual serial adapters provide a point-to-point connection from one logical partition to another, or from the Hardware Management Console (HMC) to each logical partition on the managed system. Virtual serial adapters are used primarily to establish terminal or console connections to logical partitions.

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.

The operating system console can be provided by the IVM virtual TTY, using the SOL feature.

3.2 POWER processor modes

Although, strictly speaking, not a virtualization feature, POWER modes are described in this section because they affect certain virtualization features.

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

- ▶ POWER6 compatibility mode

This execution mode is compatible with v2.05 of the Power Instruction Set Architecture (ISA). For more information, see:

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

- ▶ POWER6+ compatibility mode

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

- ▶ POWER7 mode

This is the native mode for POWER7 processors, implementing v2.06 of the Power Instruction Set Architecture. For more information, see:

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

Figure 3-2 shows how to choose the processor compatibility mode by editing the partition properties of a logical partition from the IVM.

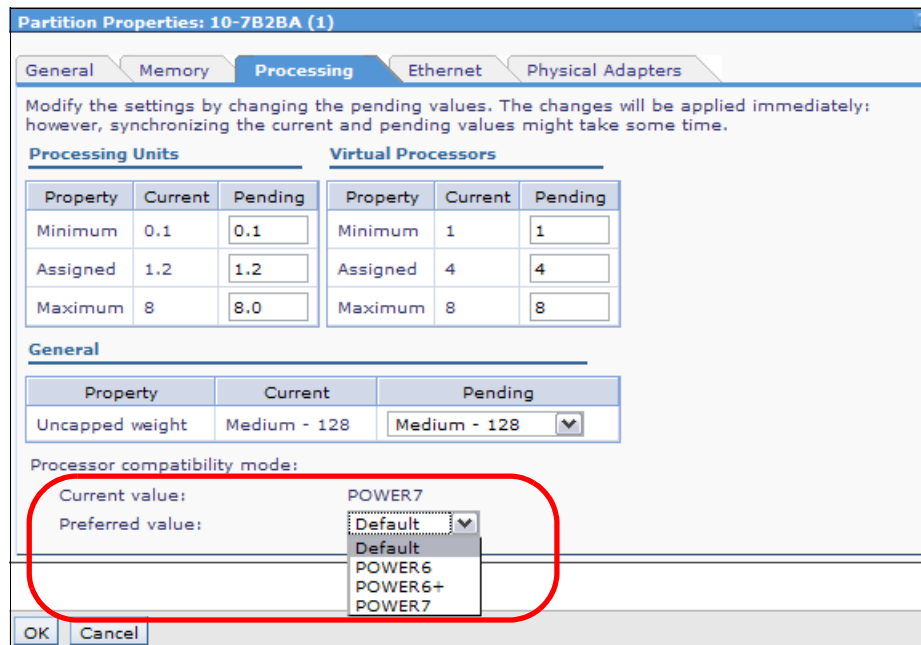


Figure 3-2 Configuring partition profile compatibility mode from IVM

Table 3-2 on page 83 lists the differences between these modes.

Table 3-2 Differences between POWER6 and POWER7 mode

| POWER6 and POWER6+ mode | POWER7 mode | Customer value |
|--|--|--|
| 2-thread SMT | 4-thread SMT | Throughput performance, processor core utilization |
| VMX (Vector Multimedia Extension or AltiVec) | VSX (Vector Scalar Extension) | High performance computing for graphic and scientific workload. |
| Affinity OFF by default | 3-tier memory, micro-partition Affinity | Improved system performance for system images spanning sockets and nodes. |
| <ul style="list-style-type: none"> ▶ Barrier Synchronization ▶ Fixed 128-byte Array; Kernel Extension Access | <ul style="list-style-type: none"> ▶ Enhanced Barrier Synchronization ▶ Variable Sized Array; User Shared Memory Access | High performance computing parallel programming synchronization facility |
| 64-core and 128-thread scaling | <ul style="list-style-type: none"> ▶ 32-core and 128-thread scaling ▶ 64-core and 256-thread scaling ▶ 256-core and 1024-thread scaling | Performance and Scalability for Large Scale-Up Single System Image Workloads (such as OLTP, ERP scale-up, WPAR consolidation). |
| EnergyScale CPU Idle | EnergyScale CPU Idle and Folding with NAP and SLEEP | Improved Energy Efficiency |

3.3 PowerVM

The PowerVM platform is the family of technologies, capabilities and offerings that deliver industry-leading virtualization on the IBM Power Systems. It is the new umbrella branding term for PowerVM (Logical Partitioning, Micro-Partitioning™, Power Hypervisor, VIOS, Live Partition Mobility, Workload Partitions, and so on). As with Advanced Power Virtualization in the past, PowerVM is a combination of hardware enablement and value-added software.

3.3.1 PowerVM Editions

This section provides information about the PowerVM Editions on POWER7 processor-based blade servers.

▶ PowerVM Express Edition

This edition is intended for evaluations, pilots, and proof of concepts, generally in single-server projects. This edition supports up to three partitions per system (VIOS, AIX, Linux, and IBM i) that share processors and I/O. It allows users to try out the Integrated Virtualization Manager (IVM) and the VIOS.

▶ PowerVM Standard Edition

This edition is intended for production deployments, and server consolidation. This edition makes the POWER7 systems an ideal platform for consolidation of AIX, Linux, and IBM i operating system applications, helping clients reduce infrastructure complexity and cost.

Offering an intuitive, Web-based interface for managing virtualization within a single blade, the IVM component of VIOS allows the small business IT manager to set up and manage logical partitions (LPARs) quickly and easily. It also enables Virtual I/O and Virtual Ethernet so that storage and communications adapters can be shared among all the LPARs running on the PS700, PS701, and PS702 Blade Servers. Ultimately, IBM micro-partitioning technology allows each processor core to be subdivided into as many as 10 virtual servers. Because the PS700, PS701, and PS702 is built with POWER7

technology, other advanced virtualization functions such as Shared Dedicated Capacity can be exploited.

► **PowerVM Enterprise Edition**

The Enterprise edition is suitable for large server deployments such as multi-server deployments and cloud infrastructure. This edition includes all the features of PowerVM Standard Edition plus a new capability called Live Partition Mobility. Live Partition Mobility allows for the movement of a running AIX or Linux partition from one POWER7 processor-based server to another with no application downtime, resulting in better system use, improved application availability, and potential energy savings. With Live Partition Mobility, planned application downtime due to regular server maintenance can be a thing of the past.

For each VIOS license ordered, an order for either the one-year (5771-VIO) or three-year (5773-VIO) Software Maintenance (SWMA) is also submitted. You must purchase a license for each active processor on the server.

Note: PowerVM Express Edition, PowerVM Standard Edition, and PowerVM Enterprise Edition are optional when running AIX or Linux. PowerVM Express Edition, PowerVM Standard Edition or PowerVM Enterprise Edition is required when running the IBM i operating system on the PS 700, PS701 and PS702 Blade Servers

Table 3-3 lists the PowerVM Edition available on each model of POWER7 processor-based blade servers with their feature code:

Table 3-3 PowerVM Edition and feature codes

| Blade Servers | Power VM Express | Power VM Standard | PowerVM Enterprise |
|---------------|------------------|-------------------|--------------------|
| PS700 | #5225 | #5227 | #5228 |
| PS701 | #5225 | #5227 | #5228 |
| PS702 | #5225 | #5227 | #5228 |

Note: It is possible to upgrade from the Express Edition to the Standard or Enterprise Edition, and from Standard to Enterprise Editions.

Table 3-4 lists the offerings of the three PowerVM editions for Power7 blades.

Table 3-4 PowerVM capabilities by edition for POWER7-based blades

| PowerVM Offerings | Express | Standard | Enterprise |
|-------------------------|--------------------|-------------|-------------|
| Micro-partitions | Yes | Yes | Yes |
| Maximum LPARs | up to 3 per server | 10 per core | 10 per core |
| Management | IVM | IVM | IVM |
| Virtual I/O Server | Yes | Yes | Yes |
| NPIV | Yes | Yes | Yes |
| Live Partition Mobility | No | No | Yes |
| Active Memory Sharing | No | No | Yes |
| LX86 | Yes | Yes | Yes |

The PowerVM Editions Web site also contains useful information:

<http://publib.boulder.ibm.com/infocenter/systems/scope/hw/index.jsp?topic=/arecu/arecukickoff.htm>

3.3.2 Logical partitions

Logical partitions (LPARs) and virtualization increase use of system resources and add a new level of configuration possibilities. This section provides details and configuration specifications about this topic. A logical partition can be regarded as a logical server, capable of booting an operating system and running a workload.

Dynamic logical partitioning

LPAR was introduced with the 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 multiple logical partitions, allowing each logical partition 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 as they are executing. IBM AIX Version 5.2, with necessary enhancements to enable dynamic LPAR, was introduced in 2002. The ability to reconfigure dynamic LPARs encourages system administrators to redefine available system resources dynamically to reach the optimum capacity for each defined dynamic LPAR.

Micro-partitioning

Virtualization of physical processors in POWER5, POWER6, and POWER7 systems introduces an abstraction layer that is implemented in POWER Hypervisor. Micro-partitioning is the ability to distribute the processing capacity of one or more physical processors among one or more logical partitions. Thus, processors are shared amongst logical partitions. Micro-partitioning technology allows you to allocate fractions of processors to a logical partition.

The POWER Hypervisor abstracts the physical processors and presents a set of virtual processors to the operating system within the micro-partitions on the system. The operating system sees only the virtual processors and dispatches runnable tasks to them in the normal course of running a workload.

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*.

When defining a shared processor partition, several options have to be defined:

- ▶ Processing Units

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.

- ▶ Cap or Uncap partition

Select whether or not the partition can access extra processing power to “fill up” its virtual processors beyond its capacity entitlement, selecting either to cap or uncap your partition. If spare processing power is available in the processor pool or other partitions are not

using their entitlement, an uncapped partition can use additional processing units if its entitlement is not enough to satisfy its application processing demand.

- ▶ **Weight**

The weight (preference) is in the case of an uncapped partition.

- ▶ **Virtual processors**

The minimum, desired, and maximum number of virtual processors. A virtual processor is a depiction or a representation of a physical processor that is presented to the operating system running in a micro-partition

The POWER Hypervisor calculates a partition's processing power based on minimum, desired, and maximum values, processing mode and on other active partitions' requirements. The actual entitlement is never smaller than the processing units desired value but can exceed that value in the case of an uncapped partition and can be 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 0.1 of a physical processor. Each physical processor can be shared by up to 10 shared processor partitions and the partition's entitlement can be incremented fractionally by as little as 0.01 of the processor. The shared processor partitions are dispatched and time-sliced on the physical processors under control of the POWER Hypervisor. The shared processor partitions are created and managed by the HMC or Integrated Virtualization Management.

Partitioning maximums on the POWER7-based blades is as follows:

- ▶ The PS700 can have four dedicated partitions or up to 40 micro-partitions
- ▶ The PS701 can have eight dedicated partitions or up to 80 micro-partitions
- ▶ The PS702 can have 16 dedicated partitions or up to 160 micro-partitions

It is important to point out that the maximums stated are supported by the hardware, but the practical limits depend on the application workload demands.

The following list details additional information about virtual processors:

- ▶ A virtual processor can be running (dispatched) either on a physical processor or as standby waiting for a physical processor to become available.
- ▶ Virtual processors do not introduce any additional abstraction level. They are only a dispatch entity. On a physical processor, virtual processors run at the same speed as the physical processor.
- ▶ Each partition's profile defines CPU entitlement, which determines how much processing power any given partition should receive. The total sum of CPU entitlement of all partitions cannot exceed the number of available physical processors in the pool.
- ▶ The number of virtual processors can be changed dynamically through a dynamic LPAR operation.

Processor 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 defines the processing mode of the logical partition.

Figure 3-3 shows a diagram of the concepts discussed in the remaining sections.

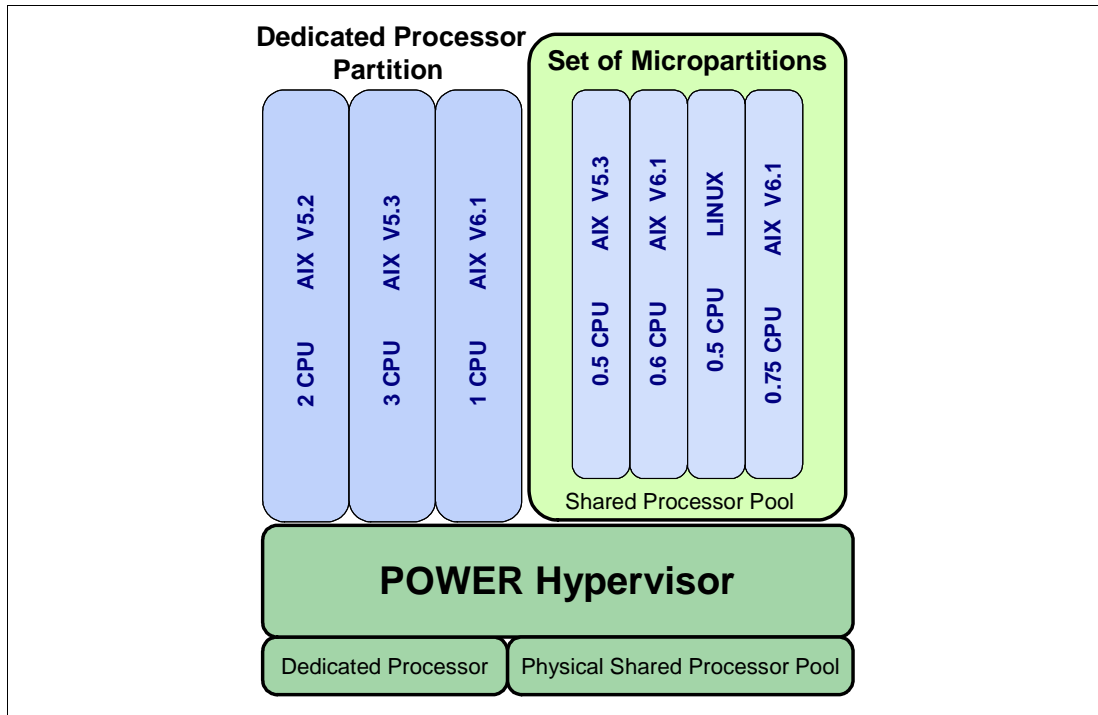


Figure 3-3 Concepts on dedicated and shared processor modes

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 as the operating system is executing (for AIX, use the `smtctl` command, and for Linux, use the `ppc64_cpu` command). If simultaneous multithreading is off, each physical processor is presented as one logical processor and thus only one thread

Shared dedicated mode

On POWER7 processor-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 use, without compromising the computing power for critical workloads in a dedicated processor.

Figure 3-4 shows how the dedicated shared processor mode can be configured.

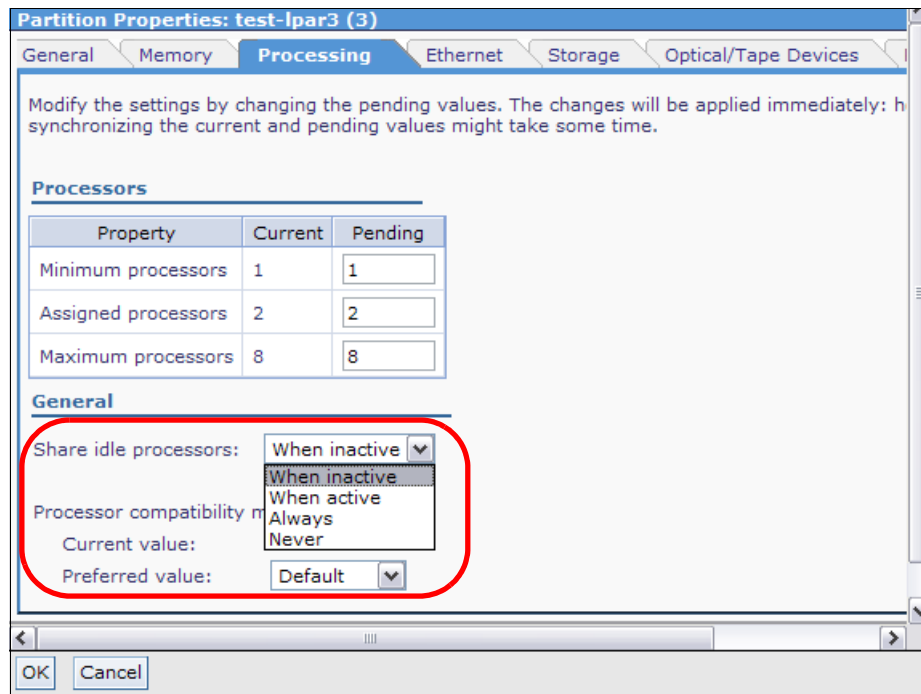


Figure 3-4 IVM console shows how to configure dedicated shared processor

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 (maximum number is 10 times the number of processing units assigned to the partition). From the POWER Hypervisor perspective, virtual processors represent dispatching objects. The POWER Hypervisor dispatches virtual processors to physical processors according to a partition's processing units entitlement. One processing unit represents one physical processor's processing capacity. At the end of the POWER Hypervisor's dispatch cycle (10 ms), all partitions receive total CPU time equal to their processing units entitlement. The logical processors are defined on top of virtual processors. Therefore, even with a virtual processor, the concept of logical processor exists and the number of logical processor depends whether the simultaneous multithreading is turned on or off.

3.3.3 VIOS

The VIOS is part of all PowerVM Editions. The Virtual I/O partition allows the sharing of physical resources between logical partitions to allow more efficient use. In this case, the VIOS owns the physical resources (SCSI, Fibre Channel, network adapters, and optical devices) and allows client partitions to share access to them, minimizing the number of physical adapters in the system. The VIOS eliminates the requirement that every partition owns a dedicated network adapter, disk adapter, and disk drive. The VIOS supports OpenSSH for secure remote logins. It also provides a firewall for limiting access by ports, network services, and IP addresses.

Figure 3-5 shows an overview of a VIOS configuration.

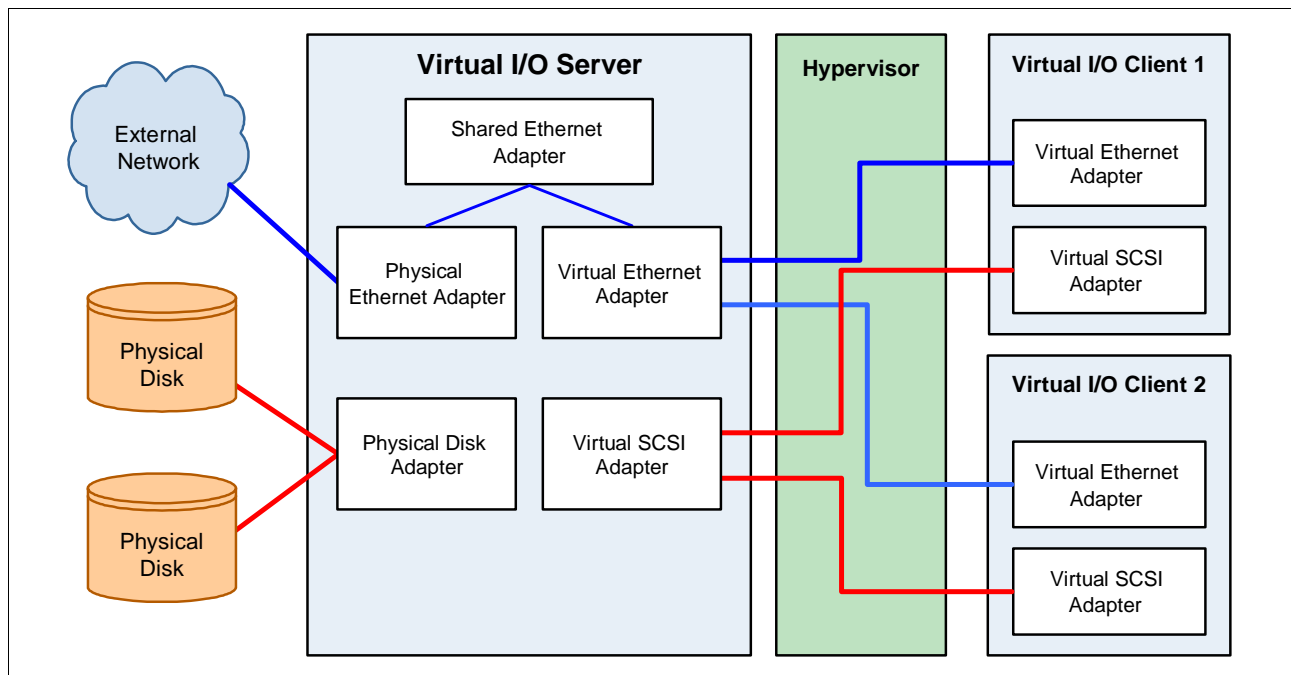


Figure 3-5 Architectural view of the VIOS

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

Installation of the VIOS partition is performed from a special system backup DVD that is provided to clients who order any PowerVM edition. This dedicated software is only for the VIOS (and IVM in case it is used) and is only supported in special VIOS partitions. Three major virtual devices are supported by the Virtual I/O Server:

- ▶ Shared Ethernet Adapter
- ▶ Virtual SCSI
- ▶ Virtual Fibre Channel adapter

The Virtual Fibre Channel adapter is used with the NPIV feature, described in 3.3.7, “N_Port ID Virtualization (NPIV)” on page 96.

Shared Ethernet Adapter

A Shared Ethernet Adapter (SEA) can be used to connect a physical Ethernet network to a virtual Ethernet network. The SEA provides this access by connecting the internal Hypervisor VLANs with the VLANs on the external switches. Because the SEA 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 SEA also provides the ability for several client partitions to share a physical adapter. Using an SEA, you can connect internal and external VLANs using a physical adapter. The SEA service can only be hosted in the VIOS, not in a general purpose AIX or Linux partition, and acts as a layer-2 network bridge to securely transport network traffic between virtual Ethernet networks (internal) and one or more (EtherChannel) physical network adapters (external). These virtual Ethernet network adapters are defined by the POWER Hypervisor on the VIOS.

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

Figure 3-6 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 VIOS that share the same physical access.

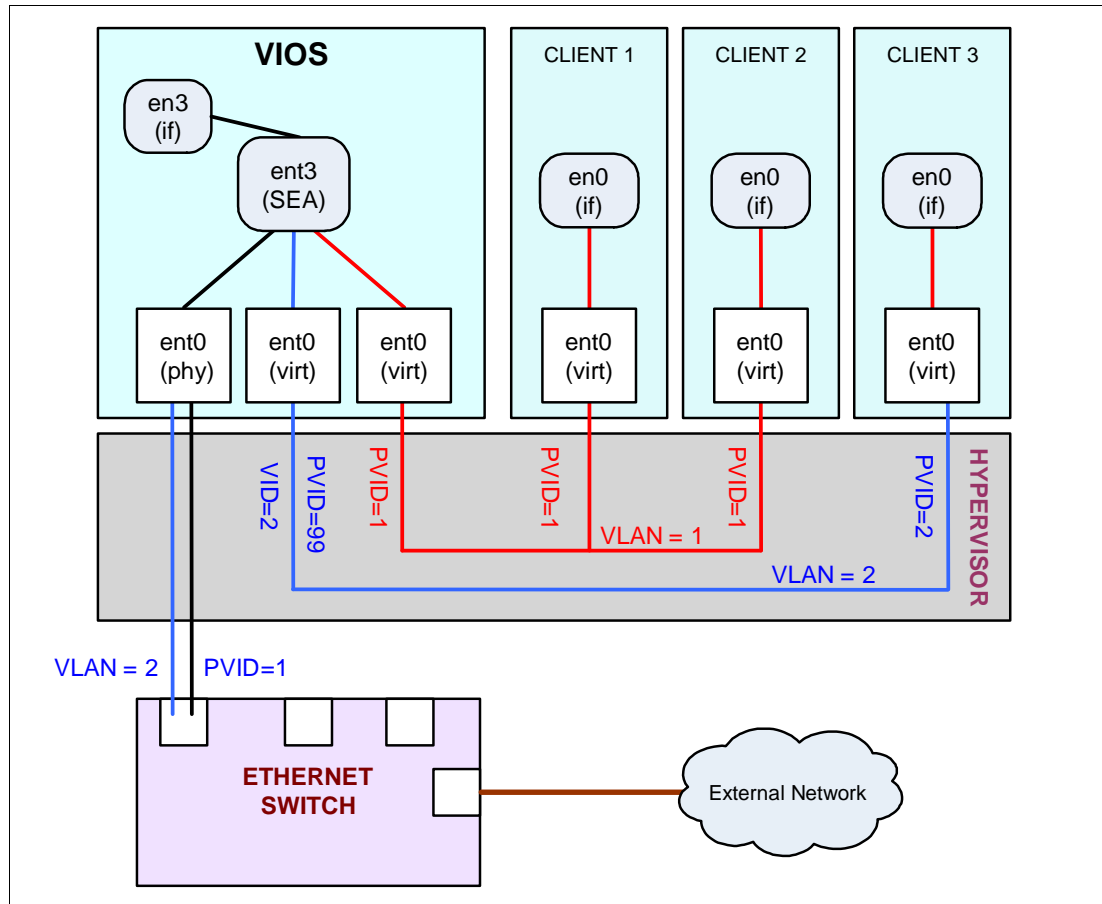


Figure 3-6 Architectural view of a SEA

A single SEA setup can have up to 16 Virtual Ethernet trunk adapters and each virtual Ethernet trunk adapter can support up to 21 VLANs (20 VIDs and 1 PVID). The number of SEAs that can be set up in a VIOS partition is limited only by the resource availability, as 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: An SEA does not need to have an IP address configured to perform the Ethernet bridging functionality. Configuring an IP address is required to access the IVM.

For a more detailed discussion about virtual networking, see the following Web page:

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

Virtual SCSI

As discussed earlier Virtual SCSI is provided by the POWER Hypervisor and allows secure communication between Client partition (AIX, Linux, or IBM i) and the VIOS. The VIOS 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 VIOS as clients.

The virtual I/O adapters (virtual SCSI server adapter and a virtual SCSI client adapter) are configured using an HMC or through the IVM on Blade Servers. The virtual SCSI server (target) adapter is responsible for executing any SCSI commands it receives. It is owned by the VIOS 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 VIOS.

Physical disks presented to the VIOS can be assigned to a client partition in a number of ways:

- ▶ The entire disk is presented to the client partition.
- ▶ The disk is divided into several logical volumes, which can be presented to a single client or multiple clients.

The logical volumes or files can be assigned to separate partitions. Therefore, virtual SCSI enables sharing of adapters as well as disk devices.

Figure 3-7 shows an example where one physical disk is divided into two logical volumes by the VIOS. 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 client partition, the disk is seen as a regular hdisk.

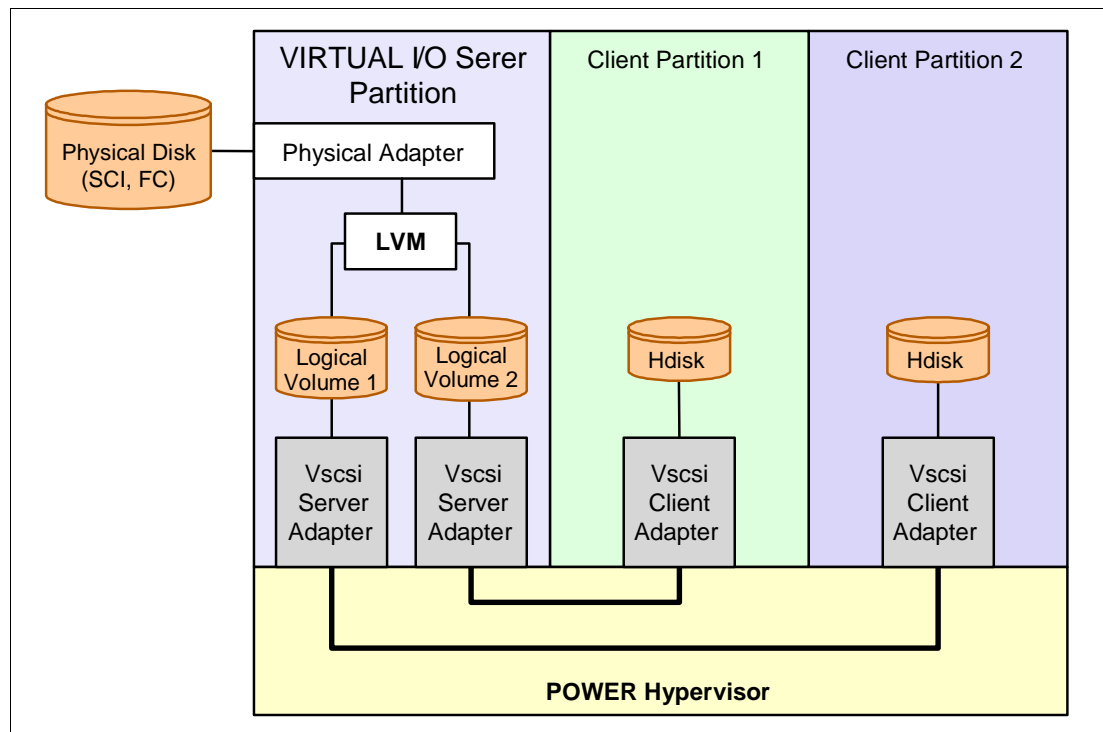


Figure 3-7 Architectural view of virtual SCSI

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 VIOS, see the following Web page:

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

VIOS functions

VIOS includes a number of features, including monitoring solutions, as follows:

- ▶ Support for Live Partition Mobility on POWER6 and POWER7 processor-based systems with the PowerVM Enterprise Edition. For more information about Live Partition Mobility, see 3.3.5, “PowerVM Live Partition Mobility” on page 93.
- ▶ 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.
- ▶ VIOS 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).
- ▶ The Workload Estimator is designed to ease the deployment of a virtualized infrastructure.
- ▶ IBM Systems Director agent and IBM Tivoli Management Tools. The IBM Systems Director Family is a suite of products consisting of a no-charge base and for-a-fee extensions. Software support is optionally available for a fee. Using IBM Systems Director and IVM, Power Systems clients can perform basic monitoring and management of their virtualized and non-virtualized environments from a single console view. Using these tools, clients can view and perform base management across IBM Systems in their enterprise. More information about IBM Systems Director can be found at the following Web page:
<http://www.ibm.com/systems/virtualization/systemsdirector/>
- ▶ Additional command-line interface (CLI) statistics in `svmon`, `vmstat`, `fcstat` and `topas` commands
- ▶ Monitoring solutions to help manage and monitor the VIOS 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, available from the following Web page:

<http://www.redbooks.ibm.com/abstracts/sg247940.html>

3.3.4 PowerVM Lx86

The PowerVM Editions hardware feature includes PowerVM Lx86. Lx86 is a dynamic, binary translator that allows Linux applications (compiled for Linux on Intel architectures) to run without change, alongside local Linux on POWER applications. Lx86 makes this possible by dynamically translating x86 instructions to POWER and caching them to enhance translation performance. In addition, Lx86 maps Linux on Intel architecture system calls to Linux on POWER architecture system calls. No modifications or recompilations of the x86 Linux applications are needed.

Lx86 creates a virtual x86 environment in which the Linux on Intel applications can run. Currently, a virtual 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. Lx86 does not run the x86 kernel on the POWER system. The Lx86 virtual environment 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.

Lx86 is included in the PowerVM Express Edition, PowerVM Standard Edition, and in the PowerVM Enterprise Edition. More information about PowerVM Lx86 can be found at the following Web page:

<http://www.ibm.com/systems/power/software/virtualization/editions/lx86/>

Note: IBM plans for PowerVM Lx86 to support POWER7 systems in second quarter 2010.

3.3.5 PowerVM Live Partition Mobility

PowerVM Live Partition Mobility allows you to move a running logical partition, including its operating system and running applications, from one system to another without disrupting the infrastructure services. The migration operation takes just a few seconds and maintains complete system transactional integrity. The migration transfers the entire system environment, including processor state, memory, attached virtual devices, and connected users.

Note: Partition Mobility is only available with the Enterprise PowerVM Edition.

A partition migration or move can occur either when a partition is powered off (inactive) or when the partition is providing service (active).

Partition mobility provides systems management flexibility and improves system availability, as follows:

- ▶ Avoid planned outages for server upgrade, hardware or firmware maintenance. Move logical partitions to another server and perform the maintenance. Live Partition Mobility can help lead to zero downtime maintenance because you can use it to work around scheduled maintenance activities.
- ▶ Meet stringent service level agreements. It allows you to proactively move the running partition and the applications from one server to another.
- ▶ Balance workload and resources. Should a key application's resource requirements peak unexpectedly to a point where there is contention for server resources, you might move it to a larger server or move other, less critical, partitions to separate servers, and use the freed-up resources to absorb the peak.
- ▶ Optimize the server. Consolidate workloads running on several small, under-used servers onto a single large server.

System requirements for Partition Mobility

Both source and destination systems must have the PowerVM Enterprise Edition license code installed. The source partition must be a virtual client and should have only 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 VIOS must be at release level 1.5.1.1 or higher has to be installed both on the source and destination systems. The MSP (mover service partition) attribute should be set to TRUE on the VIOS. The source and the target VIOS can communicate over the network. The Virtual Asynchronous Services Interface (VASI) device provides communication between the mover service partition and the POWER Hypervisor.

The minimum operating system requirement for the Partition Mobility is as follows.

- ▶ AIX 5.3 TL7 or later
- ▶ AIX 6.1 or later
- ▶ Red Hat Enterprise Linux Version 5 (RHEL5) Update 1 or later
- ▶ SUSE Linux Enterprise Server 10 (SLES 10) Service Pack 1 or later

Note: When you move an active logical partition between servers with different processor types, (such as POWER 6 and POWER7) both current and preferred compatibility modes of the logical partition must be supported by the destination server.

To migrate partitions between POWER6 and POWER7 processor-based servers. Partition Mobility can take advantage of the POWER6 Compatibility Modes that provided by POWER7 processor-based servers. On the POWER7 processor-based server, the migrated partition is then executing in POWER6 or POWER6+ Compatibility Mode.

We might 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.

The following process is an example of how to do active partition mobility from POWER6 to POWER7 Servers.

1. Set the 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. Move the logical partition to the POWER7 processor-based server. Both the current and preferred modes remain unchanged for the logical partition until you restart the logical partition.
3. Restart the logical partition on the POWER7 processor-based server. The hypervisor evaluates the configuration. Because the preferred mode is set to default and the logical partition now runs on a POWER7 processor-based server, the highest mode available is the POWER7 mode. The hypervisor determines that the most fully featured mode 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.

Now the current processor compatibility mode of the logical partition is the POWER7 mode and the logical partition runs on the POWER7 processor-based server.

Tip: The “Migration combinations of processor compatibility modes for active Partition Mobility” Web page offers presentations of the supported migrations:

<http://publib.boulder.ibm.com/infocenter/powersys/v3r1m5/topic/p7hc3/iphc3pcmco mbosact.htm>

For more information about Live Partition Mobility and how to implement it, see *IBM PowerVM Live Partition Mobility*, SG24-7460, available from the following Web page:

<http://www.redbooks.ibm.com/abstracts/sg247460.html>

3.3.6 Active Memory Sharing

Active Memory Sharing is an IBM PowerVM advanced memory virtualization technology that provides system memory virtualization capabilities to IBM Power Systems, allowing multiple partitions to share a common pool of physical memory.

The physical memory of an IBM POWER6 or POWER7 System can be assigned to multiple partitions either in a dedicated or in a shared mode. The system administrator has the capability to assign physical memory to a partition and physical memory to a pool that is shared by other partitions. A single partition can have either dedicated or shared memory.

In a dedicated memory model, the system administrator's task is to optimize available memory distribution among partitions. When a partition has performance degradation due to memory constraints and other partitions have unused memory, the administrator can allocate memory by doing a DLPAR operation.

In a shared memory model, it is the system (PowerVM Hypervisor) that automatically decides the optimal distribution of the physical memory to partitions and adjusts the memory assignment based on partition load.

Active Memory Sharing can be exploited to increase memory use on the system either by decreasing the system memory requirement or by allowing the creation of additional partitions on an existing system.

Figure 3-8 shows each logical partition can be configured to have shared memory using Active Memory Sharing or dedicated memory.

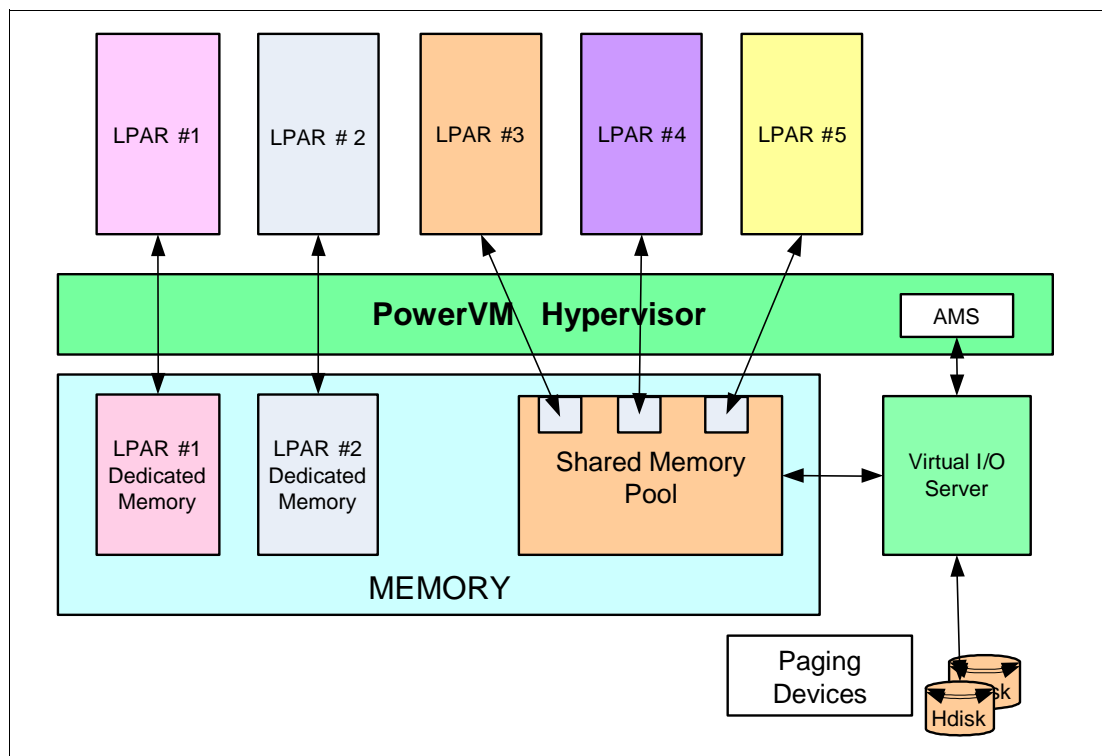


Figure 3-8 Active Memory Sharing block diagram with shared and dedicated memory

Note: Active Memory Sharing is only available with the Enterprise version of PowerVM.

System requirements for Active Memory Sharing

To use the Active Memory Sharing feature of IBM PowerVM, the following minimum system requirements for Power7 processor based blade servers must be met:

- ▶ Enterprise PowerVM Edition activation
- ▶ Virtual I/O Server Version 2.1.3
- ▶ IBM i 6.1.1 and IBM 7.1
- ▶ AIX 6.1 TL 4
- ▶ SUSE Linux Enterprise Server (SLES) 11 with SP1 or later

Figure 3-9 shows a graphical representation of a partition with different workloads at different times.

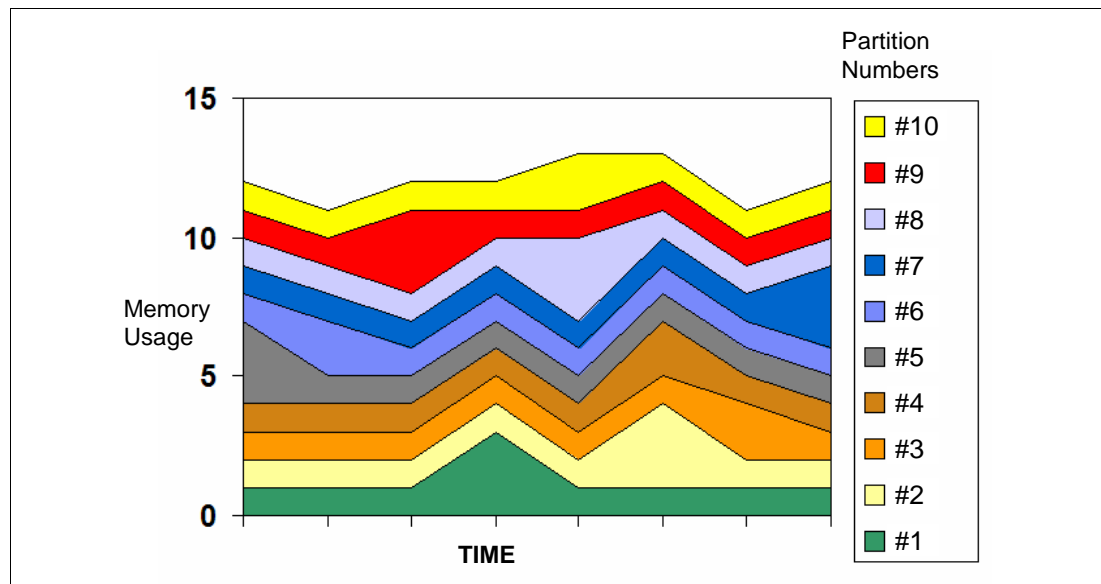


Figure 3-9 Graphical representation of Active Memory Sharing

For additional information regarding Active Memory Sharing, see *PowerVM Virtualization Active Memory Sharing*, REDP-4470, available from the following Web page:

<http://www.redbooks.ibm.com/abstracts/redp4470.html>

Note: Active Memory Expansion is not supported on POWER7 processor-based blade servers.

3.3.7 N_Port ID Virtualization (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. NPIV provides direct access to Fibre Channel adapters from multiple client partitions, simplifying the Fibre Channel SAN environment. This adapter is attached to a VIOS 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.

The following operating system levels are required for NPIV on POWER7-based blades:

- ▶ AIX V5.3 with Technology Level 12, or later
- ▶ AIX V6.1 with Technology Level 5, or later
- ▶ VIOS 2.1.3.0, or later
- ▶ IBM i 6.1.1
- ▶ IBM i 7.1
- ▶ SUSE Linux Enterprise Server 10

Table 3-5 shows NPIV compatibility matrix for AIX and Linux clients with feature codes.

Table 3-5 NPIV compatibility matrix for AIX and Linux clients

| I/O modules in the BladeCenter chassis | Expansion Card on the blade servers | | | |
|---|-------------------------------------|---|----------------------------|-------------------------------|
| | QLogic 8Gb FC CIOv (#8242) | QLogic 8Gb FC CFFh (#8271) ^a | Emulex 8Gb FC CIOv (#8240) | QLogic 10GbE CNA CFFh (#8275) |
| QLogic 4 Gb Switch Module (#3243, #3244) ^b | Yes | Yes | No | Not applicable |
| QLogic 8 Gb Switch Module (#3284) ^c | Yes | Yes | No | Not applicable |
| Brocade 4 Gb Switch Module (#3206, #3207) | No | No | Yes ^d | Not applicable |
| Brocade 8 Gb Switch Module (#5045, # 5869) | Yes | Yes | Yes | Not applicable |
| 10 GbE Passthrough Module (#5412) | Not applicable | Not applicable | Not applicable | Yes |

- a. Requires Firmware level 5.02.01 or later
- b. Requires Firmware level 6.5.0.22 or later
- c. Requires Firmware level 7.10.1.4 or later
- d. Requires the latest firmware on the Emulex CIOv adapter

Table 3-6 shows NPIV compatibility matrix for IBM i clients with feature codes

Table 3-6 NPIV Compatibility Matrix for IBM i Client

| I/O modules in the BladeCenter chassis | Expansion Card on the blade servers | | |
|---|-------------------------------------|---|----------------------------|
| | QLogic 8Gb FC CIOv (#8242) | QLogic 8Gb FC CFFh (#8271) ^a | Emulex 8Gb FC CIOv (#8240) |
| QLogic 4 Gb Switch Module (#3243, #3244) ^b | Yes ^c | Yes ^c | No |
| QLogic 8 Gb Switch Module (#3284) ^d | Yes ^c | Yes ^c | No |
| Brocade 4 gb Switch Module (#3206, #3207) | No | No | Yes ^e |
| Brocade 8 gb Switch Module (#5045, # 5869) | Yes ^c | Yes ^c | Yes ^c |
| 10 GbE Passthrough Module (#5412) | Not applicable | Not applicable | Not applicable |

- a. Requires Firmware level 5.02.01 or later
- b. Requires Firmware level 6.5.0.22 or later
- c. Virtual Tape only
- d. Requires Firmware level 7.10.1.4 or later
- e. Requires the latest firmware on the Emulex CIOv adapter

Note: NPIV support is included in PowerVM Express, Standard and Enterprise Editions.

For additional information about NPIV, see the following resources:

- ▶ *PowerVM Migration from Physical to Virtual Storage*, SG24-7825
<http://www.redbooks.ibm.com/abstracts/sg247825.html>
- ▶ *IBM PowerVM Virtualization Managing and Monitoring*, SG24-7590
<http://www.redbooks.ibm.com/abstracts/sg247590.html>

NPIV is supported in PowerVM Express, Standard, and Enterprise Editions on the IBM POWER7 processor-based systems.

3.3.8 Supported PowerVM features by operating system

Table 3-7 summarizes the PowerVM features that are supported by the operating systems and that are compatible with the POWER7 processor-based blade servers.

Table 3-7 PowerVM features supported on AIX, IBM i, and Linux operating systems

| Feature | AIX V5.3 | AIX V6.1 | IBM i 6.1.1 | SLES 10 SP3 | SLES 11 SP1 | RHEL 5.5 |
|---|------------------|------------------|------------------|------------------|-------------|----------|
| Dynamic simultaneous multithreading (SMT) | Yes ^a | Yes ^b | Yes ^c | Yes ^a | Yes | Yes |
| DLPAR I/O adapter add/remove | Yes | Yes | Yes | Yes | Yes | Yes |
| DLPAR processor add/remove | Yes | Yes | Yes | Yes | Yes | Yes |
| DLPAR memory add | Yes | Yes | Yes | Yes | Yes | Yes |
| DLPAR memory remove | Yes | Yes | Yes | No | Yes | No |
| Micro-Partitioning | Yes | Yes | Yes | Yes | Yes | Yes |
| Shared Dedicated Capacity | Yes | Yes | Yes | Yes | Yes | Yes |
| Virtual I/O Server | Yes | Yes | Yes | Yes | Yes | Yes |
| IVM | Yes | Yes | Yes | Yes | Yes | Yes |
| Virtual SCSI | Yes | Yes | Yes | Yes | Yes | Yes |
| Virtual Ethernet | Yes | Yes | Yes | Yes | Yes | Yes |
| NPIV | Yes | Yes | Yes | No | Yes | Yes |
| Live Partition Mobility | Yes | Yes | No | Yes | Yes | Yes |
| Active Memory Sharing | No | Yes | Yes | No | Yes | No |

a. Support for only two threads

b. AIX 6.1 up to TL4 SP2 supports only two threads, and supports four threads as of TL4 SP3

c. IBM i 6.1.1 and later support SMT4

Note: Most of the features listed in Table 3-7 require that IVM and VIOS be installed. Native OS does not support the functions described in Table 3-7.



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.

RAS can be described as follows:

- ▶ **Reliability**
Reliability indicates how infrequently a defect or fault in a server manifests itself.
- ▶ **Availability**
Availability indicates how infrequently the functionality of a system or application is impacted by a fault or defect.
- ▶ **Serviceability**
Serviceability indicates how well faults and their effects are communicated to users and services and how efficiently and nondisruptively the faults are repaired.

This chapter contains the following topics:

- ▶ 4.1, "Introduction" on page 100
- ▶ 4.2, "Reliability" on page 100
- ▶ 4.3, "Availability" on page 101
- ▶ 4.4, "Serviceability" on page 109
- ▶ 4.5, "Manageability" on page 120

4.1 Introduction

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, ease administrative burden, and increase system use.

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 (SER) 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 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. This includes selective redundancy to tolerate certain faults without requiring an outage or parts replacement.

The PS700, PS701, and PS702 blades are used with a BladeCenter chassis and the various components that make up the BladeCenter infrastructure. In general, the BladeCenter infrastructure RAS is outside the scope of this chapter. However, when appropriate, the BladeCenter features that enable, compliment or enhance RAS functionality on the PS700, PS701, and PS702 blades is discussed.

IBM is the only vendor that designs, manufactures, and integrates its most critical server components:

- ▶ POWER processors
- ▶ Caches
- ▶ Memory buffers
- ▶ Hub-controllers
- ▶ Clock cards
- ▶ Service processors

Design and manufacturing verification and integration, as well as field support information is used as feedback for continued improvement on the final products.

This chapter includes a manageability section describing the means to successfully manage your systems.

Several software-based availability features exist that are based on the benefits available when using AIX and IBM i as the operating system. Support of these features when using Linux varies.

4.2 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.

4.2.1 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 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. This reduces the total number of system components and reduces the total amount of heat that is generated in the design. This results 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.

4.2.2 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 correlated with relatively small increases in temperature, and POWER processor-based systems are carefully packaged to ensure adequate cooling. Critical system components, such as the POWER7 processor chips, are positioned on the blades so they receive fresh air during operation. In addition, POWER processor-based blades are installed in BladeCenter chassis that are built with redundant, variable-speed fans that can automatically increase output to compensate for increased heat in the BladeCenter chassis.

4.2.3 Redundant components and concurrent repair

High-opportunity components, or those that most affect system availability, are protected with redundancy and the ability to be repaired concurrently.

The use of redundant parts allows the system to remain operational:

- ▶ POWER7 cores include redundant bits in L1-I, L1-D, L2 caches, and L2 and L3 directories
- ▶ Redundant and hot-swap cooling in the BladeCenter chassis
- ▶ Redundant and hot-swap power supplies in the BladeCenter chassis
- ▶ Redundant integrated Ethernet ports on the blade with separate paths to independent I/O module bays in the BladeCenter
- ▶ Redundant paths for I/O expansion cards through the BladeCenter midplane to independent I/O module bays in the BladeCenter

For maximum availability, a strong recommendation is to connect power cords from the BladeCenter to two separate Power Distribution Units (PDUs) in the rack, and to connect each PDU to independent power sources.

4.3 Availability

The IBM hardware and microcode ability to monitor execution of hardware functions is generally described as the process of first-failure data capture (FFDC). This process includes predictive failure analysis. Predictive failure analysis 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 that can increase system availability, and to drive towards a high availability objective with hardware components that can perform the following functions:

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

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 POWER technology's capabilities that are focused on keeping a system environment up and running. For a specific set of functions that are focused on detecting errors before they become serious enough to stop computing work, see 4.4.1, "Detecting" on page 110.

4.3.1 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 be first 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 assigned to partitions by using a weight value or integer rating. The lowest priority partition is rated at 0 (zero) and the highest priority partition is 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.

Note: Integrated Virtualization Manager (IVM) must be installed in partition number one on the PS700, PS701, and PS702 blades when installing multiple LPARs. On IVM-managed blades, the priority value for the LPARs is changed by using the `chsyscfg` command with the `lpar_avail_priority` flag.

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.

The partition specifications for minimum, desired, and maximum capacity are 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 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, starting that partition results in an error condition that must be resolved.

4.3.2 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 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 (as the system is running) or at boot-time (IPL), depending on both 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 approach prevents the same faulty hardware from affecting system operation again, and the repair action is deferred to a more convenient, less critical time.

Persistent deallocation includes the following elements:

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

Processor instruction retry

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

Intermittent errors, often as a result of cosmic rays or other sources of radiation, are generally not repeatable.

With this function, when an error is encountered in the core, in caches and certain logic functions, the POWER7 processor automatically retries the instruction. If the source of the error was truly transient, the instruction succeeds and the system continues as before.

On IBM systems prior to POWER6, this error would have caused a checkstop.

Alternate processor retry

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

If there are available inactivated processor cores or capacity-on-demand (CoD) processor cores, the system effectively puts a CoD processor into operation after it has been determined that an activated processor is no longer operational. In this way the server remains with its total processor power.

If there are no CoD processor cores available, system-wide total processor capacity is lowered beneath the licensed number of cores.

Single processor checkstop

As in POWER6, POWER7 provides single processor check stopping for certain processor logic or command or control errors that cannot be handled by the availability enhancements mentioned in “Dynamic processor deallocation” on page 103.

This 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, errors might result on a system-wide outage.

4.3.3 Memory protection

A memory protection architecture that provides good error resilience for a relatively small L1 cache might be inadequate for protecting the much larger system main store. Therefore, a variety of protection methods are used in POWER processor-based systems to avoid uncorrectable errors in memory.

Memory protection plans must take into account many factors:

- ▶ 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. This includes the following capabilities:

- ▶ 64-byte ECC code

This innovative ECC algorithm from IBM research allows a full 8-bit device kill to be corrected dynamically. This ECC code mechanism works across DIMM pairs on a rank basis. (Depending on the size, a DIMM might have one, two, or four ranks.) With this ECC code, an entirely bad DRAM chip can be marked as bad (chip mark). After marking the DRAM as bad, the code corrects all the errors in the bad DRAM. The code can additionally mark a 2-bit symbol as bad and correct it. Providing a double-error detect or single error correct ECC or a better level of protection is additional to the detection or correction of a chipkill event.

- ▶ Hardware assisted memory scrubbing

Memory scrubbing is a method for dealing 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 transferring data between the processor and the memory uses CRC error detection with a failed operation retry mechanism and the ability to retune bus parameters dynamically when a fault occurs. In addition, the memory bus has spare capacity to substitute a spare data bit-line that is determined to be faulty.

- ▶ Chipkill

Chipkill is an enhancement that enables a system to sustain the failure of an entire DRAM chip. Chipkill spreads the bit lines from a DRAM over multiple ECC words, so that a catastrophic DRAM failure affects one bit in each word at most. The system can continue indefinitely in this state with no performance degradation until the failed DIMM can be replaced, assuming no additional single bit errors.

POWER7 memory subsystem

The POWER7 chip contains two memory controllers with four channels per memory controller. The implementation on the PS701 and PS702 uses a single memory controller per processor chip and four advanced memory buffer chips. Each memory buffer chip connects to four memory DIMMs, 16 total per processor chip. The PS700 is similar, though it only uses two memory buffer chips connecting to a total of eight DIMMs.

The bus transferring data between the processor and the memory uses CRC error detection with a failed operation retry mechanism and the ability to retune bus parameters dynamically when a fault occurs. In addition, the memory bus has spare capacity to substitute a spare data bit-line for which is determined to be faulty.

Figure 4-1 shows a POWER7 chip as implemented on a PS701 blade with its memory interface comprised of one controller and four advanced memory buffers. Advanced memory buffer chips are exclusive to IBM. They help to increase performance acting as read/write buffers. The four advanced memory buffer chips are on the system planar and support four DIMMs each.

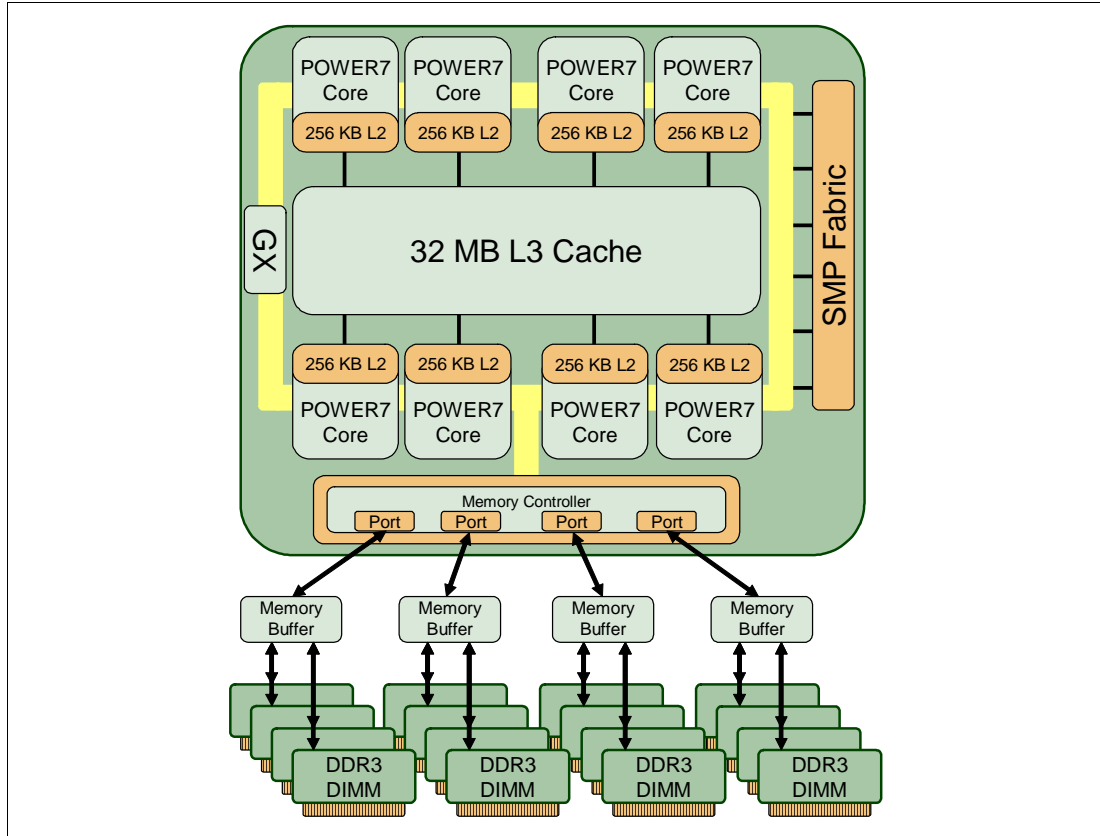


Figure 4-1 PS701 memory subsystem

Memory page deallocation

Although coincident cell errors in separate memory chips are a statistical rarity, IBM POWER 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 should be deallocated. Where possible, the operating system moves any data currently contained in that memory area to another memory area and removes the page (or pages) associated with this error from its memory map, no longer addressing these pages. The operating system performs memory page deallocation without any user intervention and is transparent to users and applications.

The POWER Hypervisor maintains a list of pages marked for deallocation during the current platform 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.

If an uncorrectable error in memory is discovered, the logical memory block that is associated with the address with the uncorrectable error is marked for deallocation by the POWER Hypervisor. This deallocation takes 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 deallocates the entire memory group associated with the error on all subsequent system reboot operations until the memory is repaired. This approach is intended to guard against future uncorrectable errors when waiting for parts replacement.

Note: Although memory page deallocation handles single cell failures, because of the sheer size of data in a data bit line, it might be inadequate for dealing with more catastrophic failures. Redundant bit steering continues to be the preferred method for dealing with these types of problems.

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.

Upon reboot, if not enough memory is available to meet minimum partition requirements, the POWER Hypervisor reduces the capacity of one or more partitions.

Depending on the configuration of the system, the IVM Electronic Service Agent™, OS Service Focal Point, or BladeCenter Advanced Management Module Service Advisor receives a notification of the failed component, and triggers a service call.

4.3.4 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's instruction and data caches are protected against intermittent errors using Processor Instruction Retry and against permanent errors by Alternate Processor Recovery. 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 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 L2 and L3.

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 does not result in a 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 would be handled through Special Uncorrectable Error handling.

L2 and L3 deleted cache lines are marked for persistent deconfiguration on subsequent system reboots until it can be replaced.

4.3.5 Special uncorrectable error handling

Although rare, an uncorrectable data error can occur in memory or a cache. IBM POWER processor-based systems attempt to limit, to the least possible disruption, the impact of an uncorrectable error using a well-defined strategy that first considers the data source. Sometimes, an uncorrectable error is temporary in nature and occurs in data that can be recovered from another repository. See the following examples:

- ▶ Data in the instruction L1 cache is never modified within the cache itself. Therefore, an uncorrectable error discovered in the cache is treated as 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 would trigger 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. Note the following information:

- ▶ If the error is irrelevant, it does not force a check stop.
- ▶ If the data is used, termination can be limited to the program or kernel, or hypervisor owning the data. Also possible is the freezing of the I/O adapters that are controlled by an I/O hub controller if data is 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 notified, and takes appropriate actions. When running AIX (since V5.2 and later) or Linux, and a process attempts to use the data, the operating system is informed of the error and might terminate, or might only terminate a specific process associated with the corrupt data. This depends on the operating system and firmware level and whether the data was associated with a kernel or non-kernel process.

Only in the case where the corrupt data is used by the POWER Hypervisor must the entire system 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 enters a freeze mode that halts normal operations. Depending on the model and type of I/O being used, the freeze might include the entire PHB chip, or a single bridge. This results in the loss of all I/O operations that use the frozen hardware until a power-on reset of the PHB is performed. The impact to partitions depends on how the I/O is configured for redundancy. In a server configured for fail-over availability, redundant adapters spanning multiple PHB chips can enable the system to recover transparently, without partition loss.

4.3.6 PCI extended error handling

IBM estimates that PCI adapters can account for a significant portion of the hardware-based errors on a large server. Although servers that rely on boot-time diagnostics can identify failing components to be replaced by hot-swap and reconfiguration, runtime errors pose a more significant problem.

PCI adapters are generally complex designs involving extensive on-board instruction processing, often on embedded microcontrollers. They tend to use industry standard-grade

components with an emphasis on product cost relative to high reliability. In certain cases, they might be more likely to encounter internal microcode errors, or many of the hardware errors described for the rest of the server.

The traditional means of handling these problems is through adapter internal error reporting and recovery techniques, in combination with operating system device driver management and diagnostics. In certain cases, an error in the adapter might cause transmission of bad data on the PCI bus itself, resulting in a hardware-detected parity error and causing a global machine-check interrupt, eventually requiring a system reboot to continue.

PCI extended error handling (EEH) enabled adapters respond to a special data packet that is generated from the affected PCI slot hardware by calling system firmware (that examines the affected bus), allowing 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.

To detect and correct PCIe bus errors, POWER7 processor-based systems use CRC detection and instruction retry correction.

4.4 Serviceability

IBM Power Systems design considers both IBM and the client's 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. Such an environment includes the following elements:

- ▶ 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 found in all POWER processor-based systems.

4.4.1 Detecting

The first and most crucial component of a solid serviceability strategy is the ability to detect errors accurately and effectively 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 is a separate microprocessor from the main instruction processing complex. The service processor provides the capabilities for the following elements:

- ▶ POWER Hypervisor (system firmware), Integrated Virtualization Manager (IVM), and BladeCenter Advanced Management Module (AMM) coordination
- ▶ Remote power control options
- ▶ Reset and boot features
- ▶ Environmental monitoring

The service processor monitors the server's built-in temperature sensors and sends this information to the BladeCenter AMM. The AMM can send instructions to the BladeCenter fans to increase rotational speed when the ambient temperature is beyond the normal operating range. Using an architected operating system interface, the service processor notifies the operating system of potential environmental problems so that the system administrator can take appropriate corrective actions before a critical failure threshold is reached.

The service processor can also post a warning and initiate an orderly system shutdown in the following circumstances:

- The operating temperature exceeds the critical level (for example, failure of air conditioning or air circulation around the system).
- The system fan speed is out of operational specification (for example, because of multiple fan failures).
- The server input voltages are out of operational specification.

The service processor can immediately shut down a system in the following circumstances:

- Temperature exceeds the critical level or if the temperature remains beyond the warning level for too long
 - Internal component temperatures reach critical levels
 - Non-redundant fan fails
 - ▶ 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 by the BladeCenter AMM, can reboot the system automatically following AC power failure.

- ▶ **Fault monitoring**

The built-in self-test (BIST) checks processor, cache, memory, and associated hardware required for proper booting of the operating system, when the system is powered on at the initial install or after a hardware configuration change (for example, an upgrade). If a non-critical error is detected or if the error occurs in a resource that can be removed from the system configuration, the booting process is designed to proceed to completion. The errors are logged in the system nonvolatile random access memory (NVRAM). When the operating system completes booting, the information is passed from the NVRAM into the system error log where it is analyzed by error log analysis (ELA) routines. Appropriate actions are taken to report the boot time error for subsequent service if required.

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:

- ▶ Continual monitoring of system operations to detect potential calculation errors.
- ▶ Attempt 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 (FFDC)

First-failure data capture (FFDC) is an error isolation technique, which 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 is captured without the need to recreate the problem or run an extended tracing or diagnostics program.

For the vast majority of faults, a good FFDC design means that the root cause is detected automatically without intervention by a service representative. Pertinent error data related to the fault is captured and saved for analysis. In hardware, FFDC data is collected from the fault isolation registers and from the associated logic. In firmware, this data consists of return codes, function calls, and so forth.

FFDC check stations are carefully positioned within the server logic and data paths to ensure potential errors can be quickly identified and accurately tracked to an FRU.

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

Figure 4-2 shows a schematic of a fault isolation register implementation.

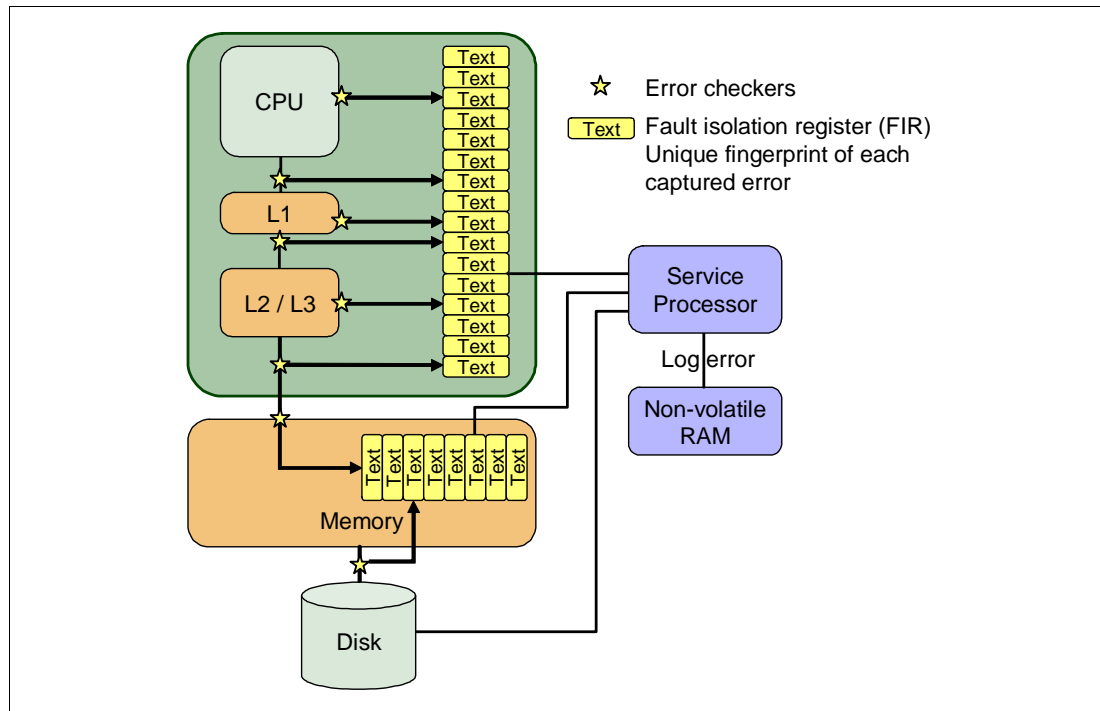


Figure 4-2 Schematic of a FIR implementation

Fault isolation

The service processor interprets error data captured by the FFDC checkers (saved in the FIRs or other firmware-related data capture methods) to determine the root cause of the error event.

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

4.4.2 Diagnosing

Using the extensive network of advanced and complementary error detection logic 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 multitier approach for system validation, starting with managed low-level diagnostics supplemented with system firmware initialization and configuration of I/O hardware, followed by OS-initiated software test routines.

Boot-time diagnostic routines include the following elements:

- ▶ 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 system processors are operational. Boot time BISTs might also find faults undetectable by a processor-based power-on self-test (POST), or through diagnostics.
- ▶ Wire-tests discover and precisely identify connection faults between components such as processors, memory, or I/O hub chips.
- ▶ Initialization of components such as ECC memory, typically by writing patterns of data and allowing the server to store valid ECC data for each location, can help isolate errors.

To minimize boot time, the system determines which of the diagnostics are required to be started to ensure correct operation based on the way the system was powered off, or through 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 checking 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. They 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 to predict when corrective actions must be automatically initiated by the system. This includes the following actions:

- ▶ Requests for a part to be replaced
- ▶ Dynamic invocation of built-in redundancy for automatic replacement of a failing part
- ▶ Dynamic deallocation of failing components so that system availability is maintained

Device drivers

In certain cases, diagnostics are best performed by operating system-specific drivers, most notably I/O devices that are owned directly by a logical partition. In these cases, the operating system device driver often works in conjunction with I/O device microcode to isolate and recover from problems. Potential problems are reported to an operating system device driver, which logs the error. I/O devices can also include specific exercisers that can be invoked by the diagnostic facilities for problem recreation if required by service procedures.

4.4.3 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 itself or the system in conjunction with a BladeCenter AMM has the ability to send an alert automatically through several methods, or contact service in the event of a critical system failure. A hardware fault also illuminates the amber system fault LED (located on the front panel of the blade) to alert the user of an internal hardware problem.

On POWER7 processor-based servers, hardware and software failures are recorded in the system log. An ELA routine analyzes the error, forwards the event to the IVM Service Focal Point (SFP) application running on the blade, 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 the BladeCenter AMM, and notifies the system administrator. After the information is logged in the SFP application and AMM event log, if the system or BladeCenter AMM are properly configured, a call-home service request is initiated. The pertinent failure data, with service parts information and part locations, is sent to an IBM Service organization. Customer 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, is 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:

- ▶ 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
- ▶ First-failure data capture 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 use this information and decide whether to contact service and support, send a notification message, or continue without an alert.

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 should only be reported once, regardless of how many logical partitions experience the potential effect of the error. The Manage Serviceable Events task, under the Service Focal Point section of the IVM user interface (Figure 4-3 on page 115), is responsible for aggregating duplicate error reports, and ensures that all errors are recorded for review and management on the single blade IVM is controlling.

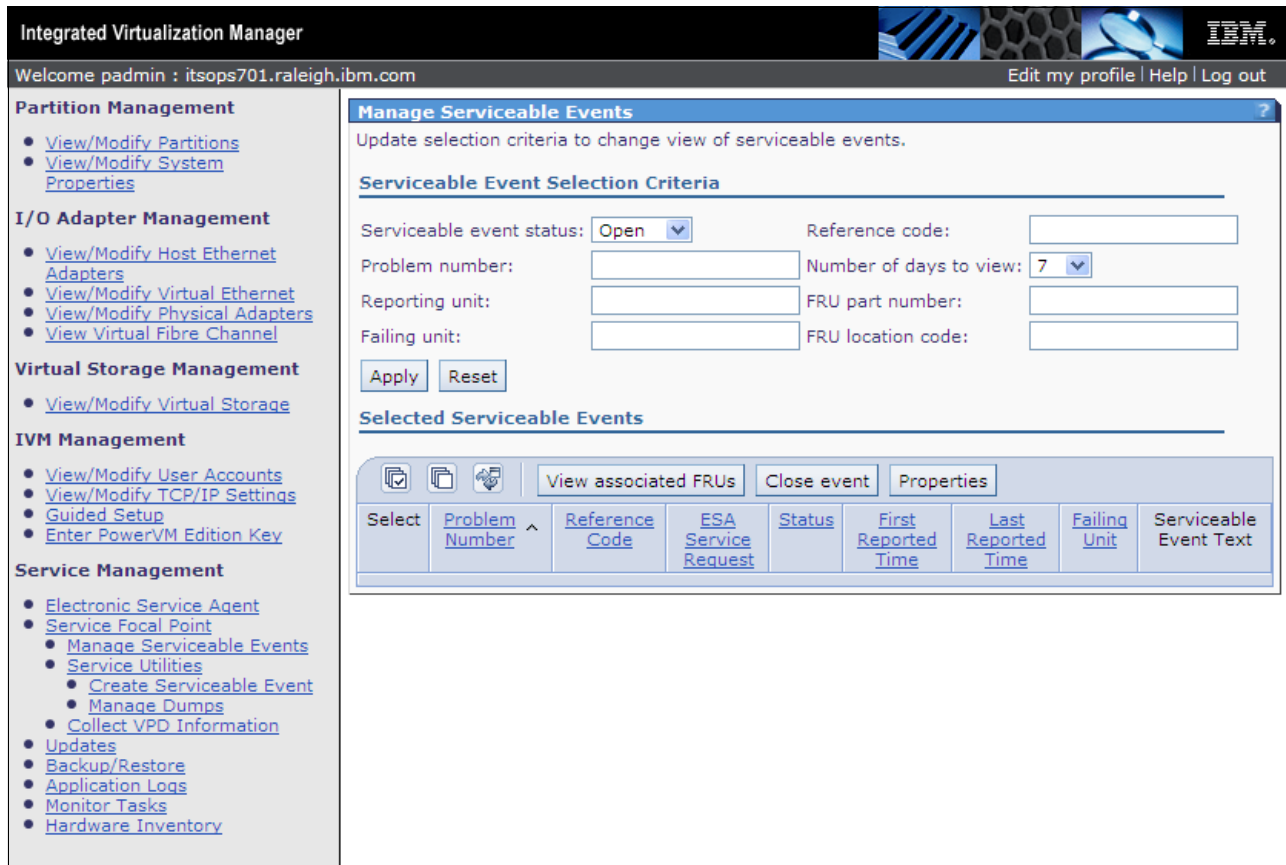


Figure 4-3 IVM Manage Serviceable Events window

The first occurrence of each failure type is recorded in Manage Serviceable Events. This task filters and maintains a history of duplicate reports from other logical partitions or the service processor. It looks at all active service event requests, analyzes the failure to ascertain the root cause and, if enabled, initiates a call for service. This methodology ensures that all platform errors are reported through at least one functional path, resulting in a single notification for a single problem.

Note: Because errors are sent to both the Service Focal Point on the blade and to the BladeCenter AMM, duplicate call home requests can be generated.

Extended error data (EED)

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 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. A manual dump creation can be done through the AMM Blade Service Data window. A service processor, platform, or partition dump can be initiated for a blade from the AMM. Service processor and platform dumps can be managed and downloaded to a workstation from the IVM Manage Dumps window, under the Service Focal Point.

4.4.4 Notifying the client

After a Power Systems server has detected, diagnosed, and reported an error to an appropriate aggregation point, it notifies the client and, if necessary, the IBM Support Organization. Depending upon the assessed severity of the error and support agreement, this could range from a simple notification to having field service personnel dispatched to the client site with the correct replacement part.

Client Notify events

When an event is important enough to report, but does not indicate the need for a repair action or the need to call 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. This includes the following examples:

- ▶ Network events such as the loss of contact over a Local Area Network (LAN)
- ▶ Environmental events such as ambient temperature warnings
- ▶ Events that need further examination by the client, but 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 they want to take further action. These events can be reported back to IBM at the client's discretion.

Call home

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

Note: Call home is used generically to indicate automatically contacting IBM service. The actual method is through an Internet connection. BladeCenter AMM and individual blades do not have modem capability.

Vital product data (VPD) and inventory management

Power Systems store 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 to assist in keeping the firmware and software on the server up-to-date.

The BladeCenter AMM also collects VPD on the individual blades and the components of the BladeCenter chassis. This information is used by support representatives to understand the complete BladeCenter/blade environment

IBM problem management database

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

4.4.5 Locating and servicing parts requiring service

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/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 screwdrivers to service the hardware components.
- ▶ Positive retention

Positive retention mechanisms assure proper connections between hardware components such as 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 used for the PS700, PS701, and PS702 blades. In the Light Path LED implementation, when a fault condition is detected on the POWER7 processor-based system, a FRU LED is illuminated, which is rolled up to the system fault LED. The Light Path system pinpoints the exact part by turning on the FRU fault LED associated with the part to be replaced. The Light Path diagnostic FRU fault LEDs can be reviewed from the AMM (as shown in Figure 4-4) or reviewed directly on the blade after removal from the BladeCenter chassis using the Light Path diagnostic switch on the blade planar.

7 - its0 redbook: Blade LED Details

Text mode

| LED Label | State | Location |
|----------------|-------|-------------|
| SAS Fault | off | Planar |
| Sys Brd Fault | off | Planar |
| Location | off | Front Panel |
| Media Tray | off | Front Panel |
| Fault | | Front Panel |
| FP Power | | Front Panel |
| Information | off | Front Panel |
| Mgmt Crd Fault | off | Planar |
| DIMM 1 Fault | off | Planar |
| DIMM 2 Fault | | Planar |
| DIMM 3 Fault | off | Planar |
| DIMM 4 Fault | | Planar |
| DIMM 5 Fault | off | Planar |
| DIMM 6 Fault | off | Planar |
| DIMM 7 Fault | off | Planar |
| DIMM 8 Fault | off | Planar |
| DIMM 9 Fault | off | Planar |
| DIMM 10 Fault | off | Planar |
| DIMM 11 Fault | off | Planar |
| DIMM 12 Fault | off | Planar |
| DIMM 13 Fault | off | Planar |
| DIMM 14 Fault | off | Planar |
| DIMM 15 Fault | off | Planar |
| DIMM 16 Fault | off | Planar |
| Battery Fault | off | Planar |
| 1Xe Fault | off | Planar |
| PCIe Fault | off | Planar |

Refresh Back

Figure 4-4 AMM blade LED details

The system can clearly identify components for replacement by using specific component-level LEDs, and can also guide the servicer directly to the component by signaling (turning on solid) the system fault LED, enclosure fault LED, and the component FRU fault LED.

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

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. The following list details several of these service labels and the purpose of each:

- ▶ Location diagrams are strategically located on the system hardware, relating information regarding the placement of hardware components. Location diagrams might include location codes, drawings of physical locations, concurrent maintenance status, or other data 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.
- ▶ The remove or replace procedure labels contain procedures often found on a cover of the system or in other spots 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 the serviceability direction of components. Certain serviceable parts (such as latches, levers, and touch points) must be pulled or pushed in a certain direction and certain order for the mechanical mechanisms to engage or disengage. Arrows generally improve the ease of serviceability.

The front panel

The front panel LEDs on the PS700, PS701, and PS702 blades indicate power status, error and informational states, disk and network activity, and physical location within a BladeCenter chassis.

Concurrent maintenance

The BladeCenter supporting infrastructure is designed with the understanding that certain components have higher intrinsic failure rates than others. The movement of fans, and power supplies make them more susceptible to wearing down or burning out. Other devices (such as I/O modules) might begin to experience wear on mechanical connectors from repeated plugging and unplugging for example, or other unexpected failure. For this reason, these devices are designed to be concurrently maintainable when properly configured.

Live Partition Mobility (LPM) provides the ability to move workload off the PS700, PS701, and PS702 blades allowing uninterrupted service when the blade or failed blade component is replaced. The use of LPM would also be used to update the blade firmware without disrupting operations.

Firmware updates

In a BladeCenter/Blade environment there are multiple areas to consider when looking at firmware updates. In most cases, BladeCenter and infrastructure components can be updated concurrently without disrupting blade operations.

POWER processor-based blades require a supported operating system running on the blade to update the system firmware. Starting with the POWER6 generation blades and continuing with the POWER7-based blades, LPM can be used to avoid the disruptive nature of blade firmware updates. Firmware updates can provide fixes to previous versions and can enable new functions. Blade system firmware typically has a prerequisite AMM firmware level.

A regular program of reviewing current firmware levels of the BladeCenter components and the blades should be in place to ensure the best availability.

Firmware updates for the AMM, I/O modules, and blades can be obtained from the IBM Fix Central web page:

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

Repair and verify system

Repair and verify (R&V) is a system used to guide a service provider 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. The following repair scenarios are covered by R&V:

- ▶ 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

R&V procedures are designed to be used both by service representative providers who are familiar with the task at hand and those who are not. Education On Demand content is placed in the procedure at the appropriate locations. Throughout the R&V 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 performed correctly.

Clients can subscribe through the subscription services to obtain the notifications on the latest updates available for service-related documentation. The latest version of the documentation is accessible through the Internet. A CD-ROM-based version is also available.

4.5 Manageability

Several functions and tools help manageability, and can allow you to efficiently and effectively manage your system.

4.5.1 Service user interfaces

The Service Interface allows support personnel or the client to communicate with the service support applications in a server and BladeCenter AMM using a console or user interface. 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 following list details the primary service interfaces:

- ▶ Light Path (for more information see “Light Path” on page 118)
- ▶ Blade LED Details (for more information see “Light Path” on page 118)
- ▶ Service Processor
- ▶ Blade front panel (for more information see “The front panel” on page 119)
- ▶ Operating system service menu
- ▶ IVM Service Management
- ▶ BladeCenter event log
- ▶ BladeCenter Service Advisor and Blade LED details

Service processor

The service processor is a controller running its own operating system. It is a component on the blade planar. The service processor operating system has specific programs and device drivers for the service processor hardware. The host interface is a processor support interface connected to the POWER processor. The service processor is always working, regardless of the main system unit's state. The system unit can be in the following states:

- ▶ Standby (power off)
- ▶ Operating, ready to start partitions
- ▶ Operating with running logical partitions

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

The two service processor Ethernet ports can be enabled by the AMM to an external network and are used to access the ASMI. The ASMI can be accessed through an HTTP server that is integrated into the service processor operating environment.

Note: The ASMI implementation in the PS700, PS701, and PS702 blades does not provide for administrator login at the current time.

Operating system service menu

The system diagnostics consist of stand-alone diagnostics that are loaded from the DVD drive in the BladeCenter media tray, and online diagnostics that are available through the operating system.

Online diagnostics, when installed, are a part of the AIX or VIOS 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. The modes are as follows:

- ▶ Service mode

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

- ▶ Concurrent mode

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

- ▶ Maintenance mode

This mode enables the checking of most system resources. Maintenance mode provides the same test coverage as service mode. The difference between the two modes is the way 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.

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

IVM Service Management (blades)

The following functions are available through the IVM Service Management:

- ▶ Electronic Service Agent or ESA (for more information see 4.5.3, “Electronic Service Agent tool” on page 124)
- ▶ Service Focal Point
 - Managing serviceable events
 - Service Utilities
 - Create Serviceable Event
 - Manage Dumps
- ▶ Collect VPD Information
- ▶ Updates (adapter capability only)
- ▶ Backup/Restore
- ▶ Application Logs
- ▶ Monitor tasks
- ▶ Hardware Inventory

BladeCenter event log

The BladeCenter event log includes entries for events that are detected by the BladeCenter unit and installed components.

The following sources can generate events that are recorded in the event log:

- ▶ Blade service processor
- ▶ BladeCenter unit
- ▶ Blade device by bay number

BladeCenter Service Advisor

The BladeCenter Service Advisor provides a method to notify a service and support representative on selected issues. When a serviceable event that has been designated as a call home event is detected, a message is written in the event log and any configured alerts are sent. The information gathered by service advisor is the same information that is available to the AMM.

4.5.2 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 is installed from a supported and running operating system on the POWER-based blade. The firmware can also be installed on a blade when booted from standalone AIX diagnostics.

Power Systems has a permanent firmware boot side, or A side, and a temporary firmware boot side, or B side. Install the new levels of firmware on the temporary side first 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 to obtain new firmware, see the following Web page:

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

For POWER based blades select the **BladeCenter** link. Figure 4-5 on page 123 is an example of the Support for IBM BladeCenter Web page.

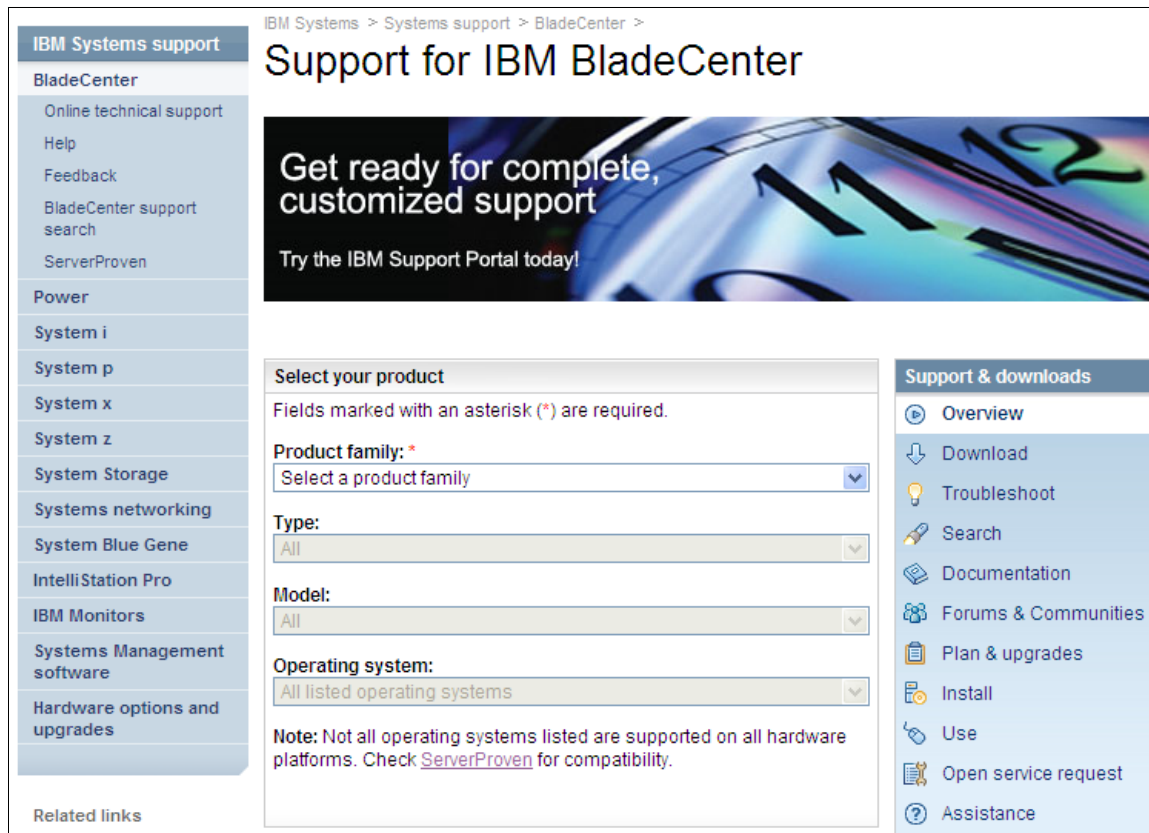


Figure 4-5 Support for IBM BladeCenter Web page

Use the “Product family” drop-down menu to select the appropriate blade type, and select **Go** to retrieve the available firmware updates and the resources for keeping your system up to date.

The current running level and boot side (A or B) of the firmware can be displayed from the AMM. The running, temporary, and permanent firmware version levels can be obtained by used of the **lsfware** command on the VIOS.

Each IBM Power Systems server has the following levels of server firmware and power subsystem firmware:

- ▶ **Installed level**
This is the level of server firmware or power subsystem firmware that has been installed and is installed into memory after the managed system is powered off and powered on. It is installed on the temporary side of system firmware.
- ▶ **Activated level**
This is the level of server firmware or power subsystem firmware that is active and running in memory.
- ▶ **Accepted level**
This is the backup level of server or power subsystem firmware. You can return to this level of server or power subsystem firmware if you decide to remove the installed level. It is installed on the permanent side of system firmware.

For POWER-based blades the installation of system firmware is always disruptive, but the effects can be mitigated by used of Live Partition Mobility.

4.5.3 Electronic Service Agent tool

Electronic Service Agent is a no-charge software tool that resides on a system to monitor events and periodically send service information to IBM support on a user-definable time table. The tool is available on VIOS, AIX, IBM i, and Linux operating systems. This tool tracks and captures service information, hardware error logs, and performance information. It automatically reports hardware error information to IBM support as long as the system is under an IBM maintenance agreement or within the IBM warranty period. Service information and performance information reporting do not require an IBM maintenance agreement or do not need to be within the IBM warranty period to be reported. Information collected by the Electronic Service Agent tool is available to IBM service support representatives to help diagnose problems.

Electronic Service Agent running on a platform is shown in Figure 4-6. Along with the IBM Electronic Services web site shown in Figure 4-7 on page 125 these tools make up IBM Electronic Services. The IBM Electronic Services web site can be accessed from the following location:

<http://www.ibm.com/support/electronic/portal>

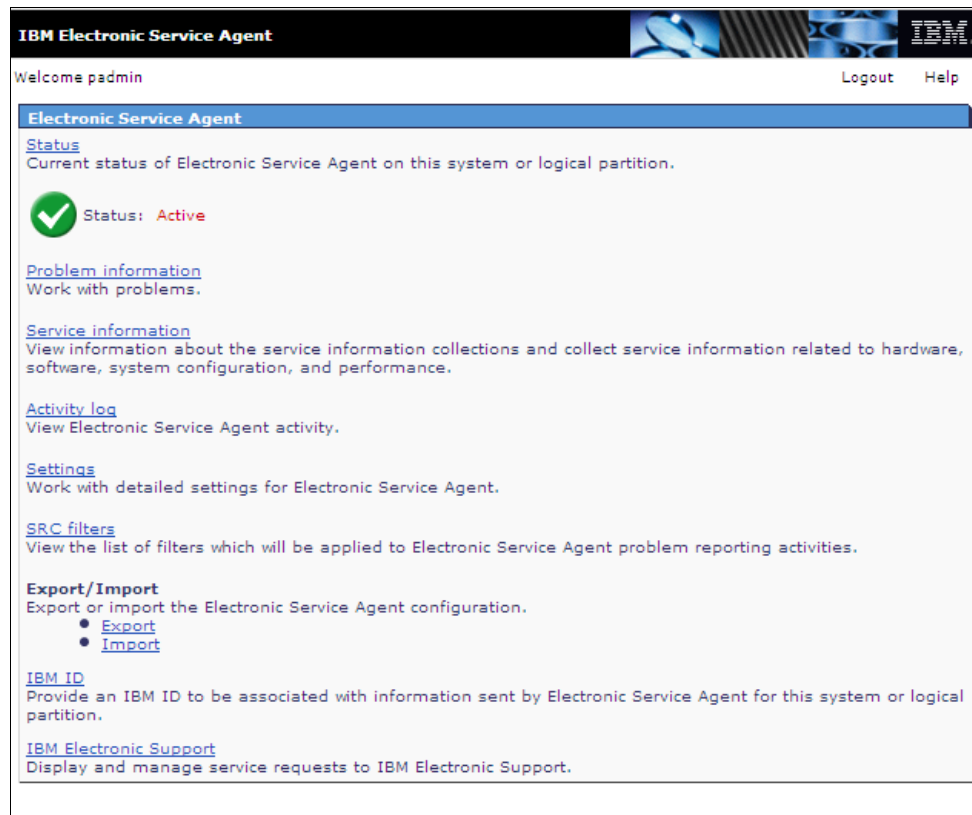


Figure 4-6 ESA welcome page



Figure 4-7 IBM Electronic Services Web site

4.5.4 BladeCenter Service Advisor

IBM BladeCenter Service Advisor comes standard in all BladeCenter chassis that have the AMM. After being configured and activated, a service event on the BladeCenter chassis can be reported to IBM Service & Support, or to a FTP/TFTP server, or to both.

IBM BladeCenter Service Advisor is built from the IBM Electronic Service Agent offering. There is no installation required for the service advisor, but it must be configured with customer information and enabled.

When a serviceable event designated as a call home event is detected, a message is written in the event log and any configured alerts sent. The information gathered by service advisor is the same information that is available if you save service data from the advanced management module Web interface.

After gathering the information, the service advisor automatically initiates a call to IBM. Upon receipt of the information, IBM returns a service request ID, which is placed in the call home activity log.

Figure 4-8 shows BladeCenter Service Advisor enabled to send alerts to both IBM Support and a FTP/TFTP server. Service advisor can be tailored through the use of the Call Home Exclusion List to specify specific call home events not to be reported.

The screenshot displays the 'Service Advisor' configuration page in the IBM BladeCenter H Advanced Management Module. The page title is 'Service Advisor' with a help icon. The main content area contains a description of the Service Advisor's function and a link to 'View Terms and Conditions'. Below this, the status for reporting to IBM Support and FTP/TFTP Server is shown as 'Enabled'. A 'Display For' dropdown menu is set to 'Both IBM Support and FTP/TFTP Server'. At the bottom, there is a table with the following data:

| Corrected | IBM Support | | FTP/TFTP Server | Event ID | Event Severity | Event Source | Date/Time |
|--------------------------|-------------|--------------|-----------------|----------|----------------|--------------|---------------------------|
| | Send | Assigned Num | | | | | |
| <input type="checkbox"/> | NO | Pending | N/A | Not Sent | 0x00016802 | Info | CHASSIS 04/07/10 17:39:38 |

Figure 4-8 BladeCenter Service Advisor

On the Event Log page of the advanced management module Web interface, you can select the Display Call Home Flag checkbox. If you select the checkbox, events are marked with a **C** for call home events and an **N** for events that are not called home. In addition, you can filter the event log view based on this setting. Figure 4-9 shows the BladeCenter event log depicting a call home event.

The screenshot shows the IBM BladeCenter H Advanced Management Module interface. The top navigation bar includes 'Welcome kcanders' and links for 'About', 'Help', and 'Logout'. The left sidebar shows the navigation menu for 'Bay 1: Lab_Services_2', with 'Event Log' selected under 'Monitors'. The main content area is titled 'Event Log' and contains the following sections:

- Options & Actions:**
 - Send service information using e-mail to report possible problems. [more...](#)
 - Download the log in Comma Separated Value (CSV) format. [more...](#)
 - Delete event log messages:
 - Monitor log state events. [more...](#)
- Filters:**
 - Note:** Hold down Ctrl to select more than one option. Hold down Shift to select a range of options. Options will not be updated pressing the "refresh" button shown on this page, only when the page is refreshed or reloaded by the browser.
 - Severity:** E - Error, W - Warning, I - Info
 - Source:** Audit, SERVPROC
 - Date:** 04/07/10
 - Serviceable:** N - Not Call Home, C - Call Home
 - Columns:** Call Home, Event ID

At the bottom, there is a table with the following data:

| Index | Sev | Source | Date / Time | Call Home | Event ID | Text |
|-------|-----|----------|-------------------|-----------|----------------------------|---------------------------------------|
| 0 | I | SERVPROC | 04/07/10 18:03:55 | C | 0x00016802 | Test Call Home generated by kcanders. |

Figure 4-9 BladeCenter event log showing call home event

Abbreviations and acronyms

| | | | |
|---------------|---|--------------|---|
| AC | alternating current | FC-IP | Fibre Channel Internet Protocol |
| AMD | Advanced Micro Devices | FDMI | Fibre Device Management Interface |
| AMM | Advanced Management Module | FFDC | first-failure data capture |
| ARP | Address Resolution Protocol | FIR | fault isolation registers |
| ASIC | application-specific integrated circuit | FRU | field replaceable unit |
| ASMI | Advanced System Management Interface | FSP | flexible service processor |
| BIOS | basic input output system | GB | gigabyte |
| BIST | Built-in self-test | HBA | host bus adapter |
| BOOTP | boot protocol | HDD | hard disk drive |
| CD-ROM | compact disc read only memory | HEA | Host Ethernet Adapter |
| CEC | central electronics complex | HMC | Hardware Management Console |
| CEE | Converged Enhanced Ethernet | HPC | high performance computing |
| CIF | Customer Installed Features | HSSM | high speed switch module |
| CLI | command-line interface | HT | Hyper-Threading |
| CMOS | complementary metal oxide semiconductor | HTTP | Hypertext Transfer Protocol |
| CNA | Converged Network Adapter | I/O | input/output |
| CPU | central processing unit | IBA | InfiniBand Architecture |
| CRC | cyclic redundancy check | IBM | International Business Machines |
| CRU | customer replaceable units | IBTA | InfiniBand Trade Association |
| CSU | Customer Set Up | ID | identifier |
| DASD | direct access storage device | IEEE | Institute of Electrical and Electronics Engineers |
| DC | domain controller | IOC | IO controllers |
| DDR | Double Data Rate | IP | Internet Protocol |
| DHCP | Dynamic Host Configuration Protocol | IPL | initial program load |
| DIMM | dual inline memory module | IPTV | Internet Protocol Television |
| DLPAR | Dynamic Logical Partition | ISA | industry standard architecture |
| DPM | Distributed Power Management | IT | information technology |
| DRAM | dynamic random access memory | ITSO | International Technical Support Organization |
| DSM | disk storage module | IVE | Integrated Virtual Ethernet |
| ECC | error checking and correcting | IVM | Integrated Virtualization Manager |
| EED | extended error data | KVM | keyboard video mouse |
| EEH | extended error handling | LAN | local area network |
| ELA | error log analysis | LDAP | Lightweight Directory Access Protocol |
| ESA | Electronic Service Agent | LED | light emitting diode |
| ETSI | European Telecommunications Standard Industry | LMB | Logical Memory Block |
| FC | Fibre Channel | LPAR | logical partitions |
| FC-AL | Fibre Channel-arbitrated loop | LPM | Live Partition Mobility |

| | | | |
|--------------|---|-------------|---|
| LUN | logical unit number | SMS | System Management Services |
| MAC | media access control | SMT | Simultaneous Multithread |
| MES | Miscellaneous Equipment Specification | SNMP | Simple Network Management Protocol |
| MPI | Message Passing Interface | SOI | silicon-on-insulator |
| MSIM | Multi-Switch Interconnect Module | SOL | Serial over LAN |
| MSP | mover service partition | SPCN | System Power Control Network |
| MTU | maximum transmission unit | SRAM | static RAM |
| NASA | National Aeronautics and Space Administration | SSA | serial storage architecture |
| NDP | Neighbor Discovery Protocol | SSD | solid state drive |
| NEBS | Network Equipment Building System | SSH | Secure Shell |
| NGN | next-generation network | SSL | Secure Sockets Layer |
| NIM | Network Installation Management | SSP | Serial SCSI Protocol |
| NL | near line | SUE | Special Uncorrectable Error |
| NPIV | N_Port ID Virtualization | SWMA | Software Maintenance Agreement |
| NVRAM | non-volatile random access memory | TB | terabyte |
| OS | operating system | TCO | total cost of ownership |
| PCI | Peripheral Component Interconnect | TL | technology level |
| PDU | power distribution unit | TPMD | thermal and power management device |
| PHB | processor host bridge | TTY | teletypewriter |
| POST | power-on self test | USB | universal serial bus |
| PS | Personal System | VASI | Virtual Asynchronous Services Interface |
| PVID | Port VLAN Identifier | VESA | Video Electronics Standards Association |
| PXE | Preboot eXecution Environment | VID | VLAN Identifier |
| QDR | quad data rate | VIOS | Virtual I/O Server |
| RAID | redundant array of independent disks | VLAN | virtual LAN |
| RAS | remote access services; row address strobe | VLP | very low profile |
| RDIMM | registered DIMM | VOIP | Voice over Internet Protocol |
| RHEL | Red Hat Enterprise Linux | VPD | vital product data |
| RSA | Remote Supervisor Adapter | WOL | Wake on LAN |
| SAN | storage area network | WWPN | World Wide Port Name |
| SAS | Serial Attached SCSI | | |
| SATA | Serial ATA | | |
| SCM | Supply Chain Management | | |
| SCSI | Small Computer System Interface | | |
| SEA | Shared Ethernet Adapter | | |
| SER | soft error | | |
| SFP | small form-factor pluggable | | |
| SLB | Segment Lookaside Buffer | | |
| SLES | SUSE Linux Enterprise Server | | |
| SMP | symmetric multiprocessing | | |

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” on page 132. Note that the documents referenced here might be available in softcopy only.

- ▶ *An Introduction to Fibre Channel over Ethernet, and Fibre Channel over Convergence Enhanced Ethernet*, REDP-4493
- ▶ *Hardware Management Console V7 Handbook*, SG24-7491
- ▶ *HPC Clusters Using InfiniBand on IBM Power Systems Servers*, SG24-7767
- ▶ *IBM BladeCenter JS12 and JS22 Implementation Guide*, SG24-7655
- ▶ *IBM BladeCenter JS23 and JS43 Implementation Guide*, SG24-7740
- ▶ *IBM BladeCenter Products and Technology*, SG24-7523
- ▶ *IBM PowerVM Live Partition Mobility*, SG24-7460
- ▶ *IBM PowerVM Virtualization Managing and Monitoring*, SG24-7590
- ▶ *Integrated Virtual Ethernet Adapter Technical Overview and Introduction*, REDP-4340
- ▶ *Integrated Virtualization Manager on IBM System p5*, REDP-4061
- ▶ *PowerVM Migration from Physical to Virtual Storage*, SG24-7825
- ▶ *PowerVM Virtualization Active Memory Sharing*, REDP-4470
- ▶ *PowerVM Virtualization on IBM System p: Introduction and Configuration Fourth Edition*, SG24-7940

Other publications

These publications are also relevant as further information sources, available from <http://http://ibm.com/systems/support> (click BladeCenter):

- ▶ *IBM BladeCenter PS700 Installation and User's Guide*
- ▶ *IBM BladeCenter PS700 Problem Determination and Service Guide*
- ▶ *IBM BladeCenter PS701 and PS702 Installation and User's Guide*
- ▶ *IBM BladeCenter PS701 and PS702 Problem Determination and Service Guide*

Online resources

This Web site is also relevant as further information sources:

- ▶ IBM BladeCenter PS700, PS701, and PS702 Express home page
<http://ibm.com/systems/bladecenter/hardware/servers/ps700series>

How to get Redbooks

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

ibm.com/redbooks

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services



IBM BladeCenter PS700, PS701, and PS702 Technical Overview and Introduction



Features the POWER7 processor providing advanced multi-core technology

Details the follow-on to the BladeCenter JS23 and JS43 servers

Includes product information and features

The IBM BladeCenter PS700, PS701, and PS702 are premier blades for 64-bit applications. They are designed to minimize complexity, improve efficiency, automate processes, reduce energy consumption, and scale easily. These blade servers are based on the IBM POWER7 processor and support AIX, IBM i, and Linux operating systems. Their ability to coexist in the same chassis with other IBM BladeCenter blades servers enhances the ability to deliver rapid return of investment demanded by clients and businesses.

This IBM Redpaper is a comprehensive guide covering the IBM BladeCenter PS700, PS701, and PS702 servers. The goal of this paper is to introduce the offerings and their prominent features and functions.

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