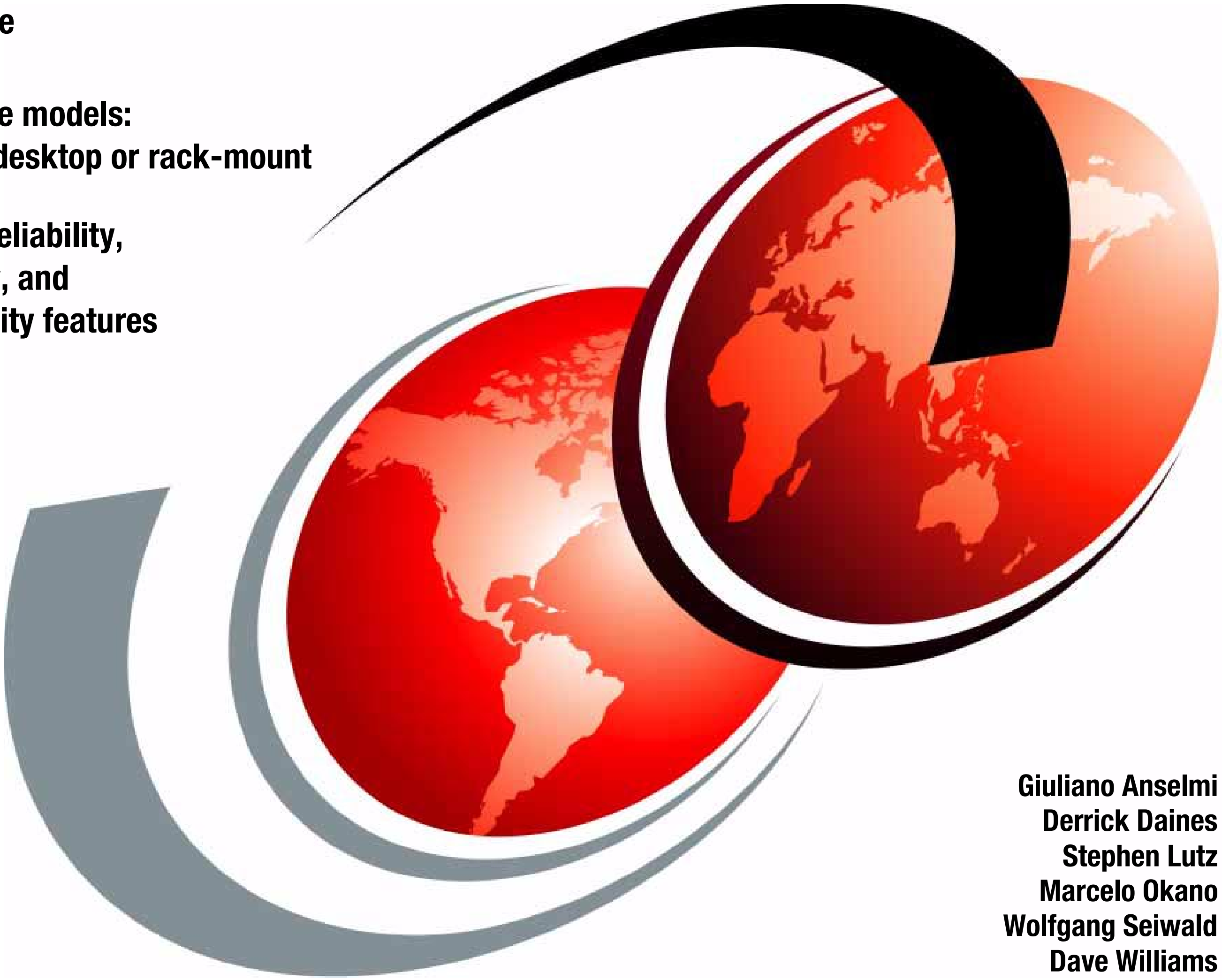
IBM eserver

pSeries 630 Models 6C4 and 6E4 Technical Overview and Introduction

Logical partitionable, I/O drawer expandable

Two unique models: Deskside/desktop or rack-mount

High-end reliability, availability, and serviceability features



Giuliano Anselmi
Derrick Daines
Stephen Lutz
Marcelo Okano
Wolfgang Seiwald
Dave Williams

Redpaper

ibm.com/redbooks

International Technical Support Organization

IBM @server pSeries 630 Models 6C4 and 6E4 Technical Overview and Introduction

April 2003

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

Sixth Edition (April 2003)

This edition applies to the IBM @server™ pSeries™ 630 Models 6C4 and 6E4 and AIX 5L™ Version 5.2, product number 5765-E62.

Contents

Notices

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

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

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

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

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

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

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

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

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

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

Trademarks

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

AIX®
AIX 5L™
AS/400®
Chipkill™
ClusterProven®
Electronic Service Agent™
Enterprise Storage Server™
@server™
IBM®
IBMLink™
iSeries™
Lotus®
Notes®
Perform™
POWER4™
POWER4+™
PowerPC®
pSeries™
Redbooks™
Redbooks(logo)™
RS/6000®
Service Director™
SP™
Word Pro®

The following terms are trademarks of other companies:

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

ActionMedia, LANDesk, MMX, Pentium and ProShare are trademarks of Intel Corporation in the United States, other countries, or both.

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

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

C-bus is a trademark of Corollary, Inc. in the United States, other countries, or both.

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

SET, SET Secure Electronic Transaction, and the SET Logo are trademarks owned by SET Secure Electronic Transaction LLC.

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

Preface

This document is a comprehensive guide covering the IBM @server pSeries 630 Models 6C4 and 6E4 entry servers. Major hardware offerings are introduced and their prominent functions discussed.

Professionals wishing to acquire a better understanding of IBM @server pSeries products may consider reading this document. The intended audience includes:

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

This document expands the current set of IBM @server pSeries documentation by providing a desktop reference that offers a detailed technical description of the pSeries 630 Models 6C4 and 6E4.

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

The team that wrote this Redpaper

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

Giuliano Anselmi is a pSeries systems product engineer in the EMEA Mid-range Field Support Organization in Rome, Italy, supporting the Web Server Sales Organization in EMEA, IBM sales, IBM Business Partners, and technical support organizations. He is a resolution assistant resource for critical situations escalated and related to IBM RS/6000® and pSeries Systems for the pSeries Customer Satisfaction Project Office, and IBM Dublin Enterprise Servers manufacturing plant. He works for IBM and has been devoted to RS/6000 and pSeries systems for 10 years.

Derrick Daines is the pSeries specialist for the Bedfont office in the UK. He has 13 years of experience working on pSeries products. He has worked at IBM for 34 years on products, including the pSeries 690. He provides support to other engineers within his own branch, other branches within the region, and he also provides country support during weekends.

Stephen Lutz is an IT Specialist for pre-sales technical support for IBM @server pSeries and RS/6000, part of the Web Server Sales Organization in Stuttgart, Germany. He holds a degree in Computer Science from the Fachhochschule Karlsruhe - University of Technology and is an IBM Certified Advanced Technical Expert. Stephen is a member of the High-End Technology Focus Group, supporting IBM sales, IBM Business Partners, and customers with pre-sales consultation and implementation of client/server environments.

Marcelo Okano is a Senior IT Specialist for pre-sales technical support for IBM @server pSeries and RS/6000 in Brazil. He has 10 years of experience working on pSeries products. He holds a degree in Mathematics from the Faculdade de Filosofia, Ciencias e Letras de Santo Andre. He has worked at IBM for four years. His areas of expertise include pSeries and RS/6000 systems, RS/6000 SP and Cluster 1600 systems, HACMP, HAGEO, Linux, performance tuning, and AIX.

Wolfgang Seiwald is a presales technical support specialist at the IBM office in Salzburg, Austria. He has worked at IBM for three years. He holds a Diplomingenieur degree in Telematik from the Technical University of Graz. The main focus of his work lies in the areas of the IBM @server systems and the IBM AIX operating system.

Dave Williams is a consulting IT specialist in the UK Advanced Technical Support group based in Bedfont. He has 25 years of experience at IBM, the last 18 of them being with IBM AIX products. His areas of expertise include pSeries 690 and e-business solutions.

The project that created this document was managed by:

Scott Vetter of the International Technical Support Organization, Austin Center

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

IBM USA
Doug Bossen, Luke Browning, Brad Frey, Ghadir Gholami, John Hilburn, Tenley Jackson, Jim McGaughan, Andy McLaughlin, Bill Mihaltse, John O'Quin, Jan Palmer, Jim Reidy, Craig Shempert, Alan Standridge, Ven Tenjarla, Ed Toutant, Sterling Weaver, Dave Willoughby

IBM Germany
Volker Haug, Gregor Linzmeier, Ralf Vollmer

IBM UK
Nigel Griffiths

Comments welcome

Your comments are important to us.

We want our papers to be as helpful as possible. Send us your comments about this in one of the following ways:

- Send your comments in an Internet note to:

  `sbv@us.ibm.com`

- Mail your comments to:

  IBM Corporation, International Technical Support Organization
  ATTN: Scott Vetter
  Dept. JN9B Building 003 Internal Zip 2834
  11400 Burnet Road
  Austin, Texas 78758-3493, USA

1

General description

The IBM @server pSeries 630 Models 6C4 and 6E4 (referred to hereafter as the Model 6C4 and Model 6E4 or p630 when discussing both models) are designed for customers looking for cost-effective, high-performance, space-efficient servers that make use of IBM technology, first used in the high-end pSeries 690. These systems use the 64-bit, copper/SOI-based, POWER4 and POWER4+ microprocessors, packaged as 1- and 2-way cards.

The Models 6C4 and 6E4 are members of the 64-bit family of symmetric multiprocessing (SMP) UNIX servers from IBM. The Model 6C4 (product number 7028-6C4) is a 4 EIA[1] (4U) 19-inch rack-mounted server, while the Model 6E4 (product number 7028-6E4) is a deskside/desktop server. With a maximum of two processor cards, the Models 6C4 and 6E4 can be configured into 1-, 2-, or 4-way systems. Each processor card is packaged together with up to 16 GB of memory per card into a processor book (a sealed unit that protects the components in a rigid structure designed for higher reliability). Total system memory can range from 1 GB up to 32 GB on a 4-way system based on the currently available DIMMs.

The availability of dynamic logical partitioning (LPAR) (2.8, “LPAR” on page 27) and cluster support (3.3, “IBM @server Cluster 1600 and SP switch attachment” on page 40) enhances the already exceptional value of these models.

The Models 6C4 and 6E4 include six hot-plug PCI-X slots and an integrated Single Ended SCSI controller (when equipped with POWER4+ processors), dual integrated Ultra3 SCSI controllers, dual 10/100 Mbps integrated Ethernet controllers, and four front-accessible disk bays supporting hot-swappable disks. These disk bays can accommodate up to 587.2 GB of disk storage using 146.8 GB Ultra3 SCSI disk drives. Two media bays are used for a CD-ROM, DVD-RAM, DVD-ROM, or another optional media device, such as a tape or diskette drive. The Converged Service Processor[2] (CSP) functionality, including system power control, is also integrated, along with the native I/O functions such as serial ports, keyboard, and mouse.

The Models 6C4 and 6E4 contain an enhanced I/O subsystem with the implementation of the remote I/O (RIO) interconnect and PCI-X bus protocols. The Model 6C4 can support up to two high-density 7311-D20 I/O drawers to provide additional PCI-X slots and disk drive bays.

[1] One Electronic Industries Association Unit (EIA) is 44.45 mm (1.75 inch).

[2] Since the release of 7025 Model F80, 7026 Model H80, and M80, the RS/6000 (pSeries) Service Processor design converged to AS/400 (iSeries) Service Processor design.

A fully configured POWER4+ system with two I/O drawers has 20 PCI-X slots (18 on POWER4 systems) and 4.1 TB of disk space.

Additional reliability and availability features include optionally redundant hot-plug cooling fans and power supplies. Along with these hot-plug components, these systems are designed to provide an extensive set of reliability, availability, and serviceability (RAS) features that include improved fault isolation, recovery from errors without stopping the system, avoidance of recurring failures, and predictive failure analysis. See 3.4, “Reliability, availability, and serviceability (RAS) features” on page 38, for more information on RAS features.

The pSeries 630 Model 6C4 is an ideal replacement for the pSeries 640 Model B80 and the pSeries 630 Model 6E4 for the RS/6000 Model 270.

1.1 Physical packages

The following sections discuss the major physical attributes found on the p630 servers and the optional I/O drawer.

1.1.1 Models 6C4 and 6E4

Figure 1-1 shows an inside view and external view of the Models 6C4 and 6E4 in order for you to be familiar with the locations of all the various parts and devices.

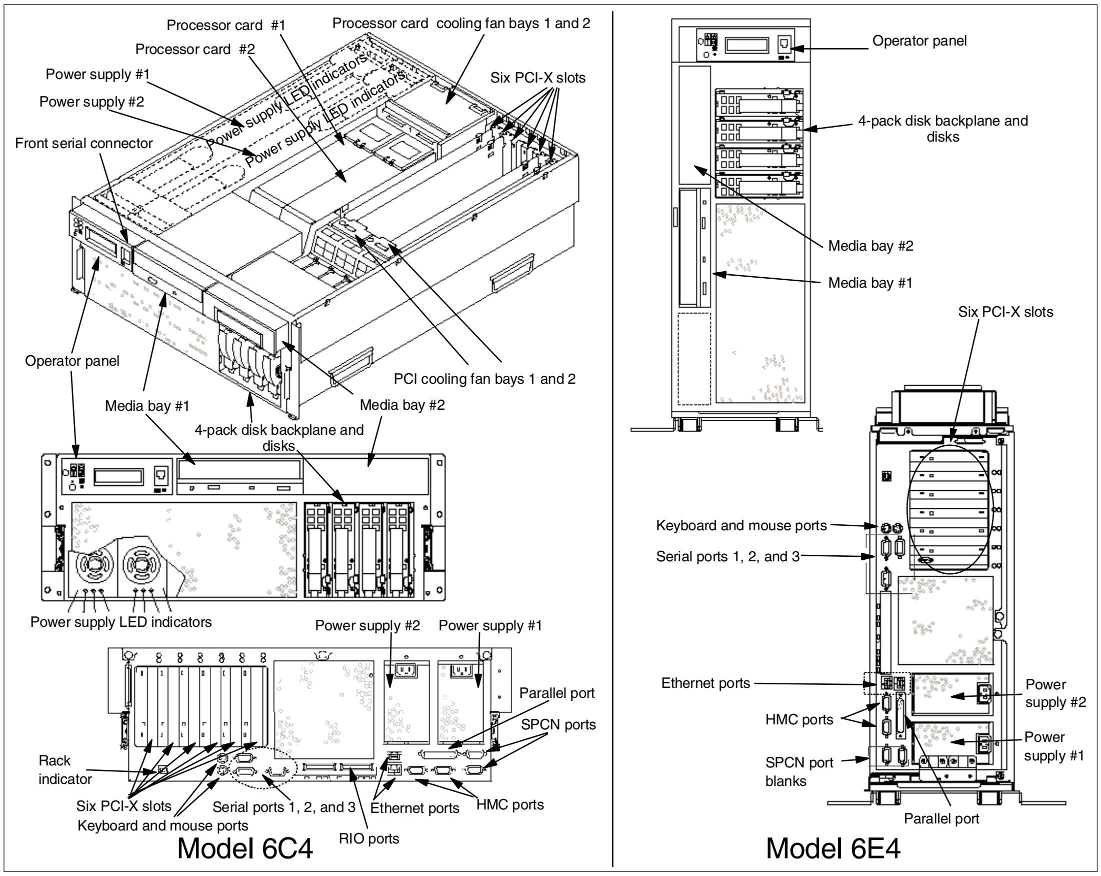


Figure 1-1 Views of Models 6C4 and 6E4 (POWER4+ system with six PCI-X slots)

The Model 6C4 can be mounted in existing 7014 Model T00 and 7014 Model T42 racks.

The physical characteristics of the Models 6C4 and 6E4 are provided in Table 1-1.

Table 1-1 Models 6C4 and 6E4 physical characteristics

Dimensions	**Rack (Model 6C4)**	**Tower (Model 6E4)**
Height	176 mm (6.9 in.) 4 EIA Units	544 mm (21.42 in.)
Width	448 mm (17.6 in.)	308 mm (12.13 in.) 191 mm without legs.
Depth	816 mm (32.13 in.) - Includes 145 mm for cable management arm	789.0 mm (31 in.) - Includes 70 mm for rear acoustic cover
Weight		
Minimum configuration	32.0 Kg (70.4 lbs.)	36.0 Kg (79.2 lbs.)
Maximum configuration	47.3 Kg (104.0 lbs.)	51.0 Kg (112.2 lbs.)

1.1.2 7311-D20

The rack-mount 7311-D20 I/O drawer is 4U. The I/O drawer contains seven PCI-X slots and can be equipped with up to 12 disks. The physical characteristics of the 7311-D20 I/O drawer are provided in Table 1-2.

Table 1-2 7311-D20 physical characteristics

Dimensions	
Height	172.8 mm (6.8 in.)
Width	431.8 mm (17.0 in.)
Depth	609.6 mm (24.0 in.)

Figure 1-2 shows the front and rear views of the 7311-D20 I/O drawer.

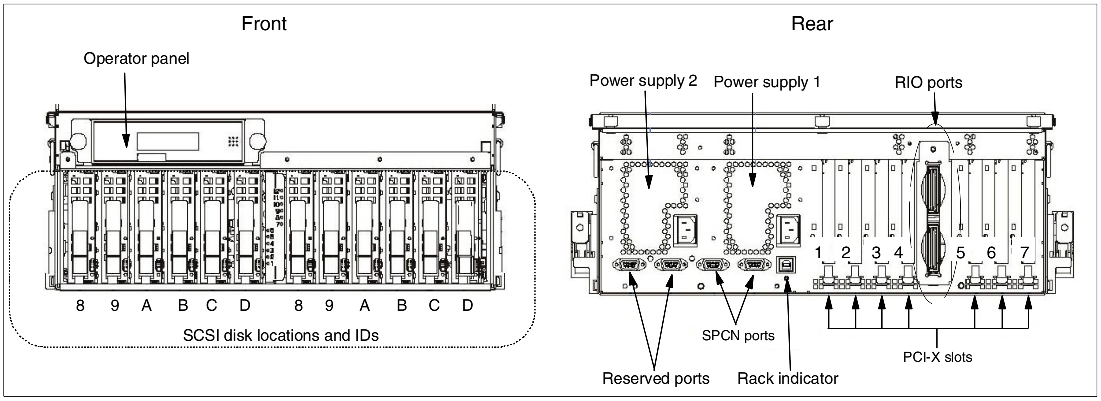


Figure 1-2 Views of 7311-D20 I/O drawer

1.2 Minimum and optional features

The p630 delivers a cost-efficient growth path for the future through the capabilities covered in this section.

Processor and memory

For the processor and memory:

- 1-, 2-, and 4-way SMP system (CUoD[3] is not offered on p630 systems)
  - 1- and 2-way processor books with 1.0 GHz POWER4 microprocessors with 32 MB of L3 cache per processor card.
  - 1- and 2-way processor books with 1.2 GHz or 1.45 GHz POWER4+ microprocessors with 8 MB of L3 cache per processor card.
  - Processor configuration guidelines:
    - The p630 supports a 2-way configuration using a single 2-way processor card (FC 5132, 5134, or 5127) or two 1-way processor cards (FC 5131, 5133, or 5126).
    - A 2-way configuration using two 1-way processor cards requires firmware Level RR 030324. This firmware can be downloaded at:

      https://techsupport.services.ibm.com/server/mdownload
    - 3-way configurations are not supported.
    - POWER4 and POWER4+ cards cannot be intermixed.
    - Processors with different speeds cannot be mixed.
    - 1.2 GHz POWER4+ processors (FC 5133 or 5134) require a 6-slot riser card (FC 9556 Initial Order or FC 6556 MES) and a RIO enabled planar (FC 6575 or FC 9575).
    - 1.45 GHz POWER4+ processors (FC 5126 or FC 5127) require a 6-slot riser card (FC 9556 Initial Order or FC 6556 MES), a RIO enabled planar (FC 6575 or FC 9575) and a Redundant Cooling Option (FC 6557).
  - Processor upgrades:
    - Processor upgrades are obtained through feature conversions.
    - A 1-way to 2-way configuration upgrade requires a installation of a second 1-way processor card or one feature conversion from 1-way to 2-way processors card.
    - A 2-way configuration using two 1-way processor cards to a 4-way configuration upgrade requires two feature conversions from 1-way to 2-way processors card.
- 1 GB to 32 GB ECC DDR SDRAM memory (based on DIMMs at time of publication)
  - Memory DIMMs plug into the processor book (eight DIMM slots per card).
  - DIMMs must be populated in quads (four DIMMs). A memory feature consists of a quad. Additional quads may consist of any memory feature code (memory size). The following are the FCs available at the time this publication was written:
    - FC 4451, 1024 MB (4 x 256 MB), DIMMS, 208-pin 8 ns DDR SDRAM.
    - FC 4452, 2048 MB (4 x 512 MB), DIMMS, 208-pin 8 ns DDR SDRAM.
    - FC 4453, 4096 MB (4 x 1024 MB), DIMMS, 208-pin 8 ns stacked DDR SDRAM.
    - FC 4454, 8192 MB (4 x 2048 MB), DIMMS, 208-pin 8 ns stacked DDR SDRAM.
  - A system with a single processor card (1- or 2-way) may have a maximum of 16 GB of memory based on the maximum memory feature available (FC 4454 4 x 2048 MB).
  - A 2-way system using two 1-way processor cards (FC 5132, 5134 or 5127) provide 16 DIMM slots and each processor card has eight DIMM slots and can be populated using an 8 GB memory option (FC 4454 4 X 2058 MB) to a maximum of 32 GB of memory.

[3] Capacity Upgrade on Demand

- Memory configuration guidelines:
  - The placement order is from the lowest to highest value quads (4451, 4452, 4453, or 4454).
  - The memory in the processor is placed in the first quad and then the second quad.
  - If two processors cards are installed, first fill the first quads on both processors and then the second quads.

Hot-swappable disk drives inside the Model 6C4 and 6E4

The Models 6C4 and 6E4 contain four hot-swappable disk drive bays. Each bay may contain one of the following features:

- 18.2 GB Ultra3 10K RPM (FC 3157)
- 36.4 GB Ultra3 10K RPM (FC 3158)
- 36.4 GB Ultra3 15K RPM (FC 3277)
- 73.4 GB Ultra3 10K RPM (FC 3159)
- 73.4 GB Ultra3 15K RPM (FC 3278)
- 146.8 GB Ultra3 10K RPM (FC 3275)

Media bays

Two media bays are available (one CD-ROM, DVD-RAM, or DVD-ROM is required). The IDE DVD-ROM or SCSI DVD-RAM can read CD-ROM installation media. Table 1-3 lists the available devices for Models 6C4 and 6E4:

- Media bay 1 can accommodate IDE or SCSI devices.
- Media bay 2 can only accommodate SCSI devices or the diskette drive.

Table 1-3 Available media devices for Models 6C4 and 6E4

Feature Code	Description	Media bay 1	Media bay 2
2605	Diskette drive		X
2623	DVD-RAM drive 4.7 GB capacity	X	X
2633	650 MB IDE 48x CD-ROM drive	X	
2634	4.7 GB IDE 16x/48x DVD-ROM drive	X	
6120	80/160 GB VXA tape drive		X
6134	8 mm 60/150 GB tape drive		X
6158	4 mm 20/40 GB tape drive		X

PCI-X slots and integrated adapters

The following devices are included in the Model 6C4 and 6E4:

- Six hot-plug PCI-X slots, 64-bit, 133 MHz, 3.3 volt on POWER4+ processor models, four hot-plug PCI-X slots on POWER4 processor models.

  For more information on PCI-X, see 2.4, “PCI-X slots and adapters” on page 18.
- Integrated ports.
  - Two 10/100 Ethernet (IEEE 802.3 compliant).

- Two Ultra3 SCSI (one external Ultra3 SCSI (with VHDCI[4] mini 68-pin port), and one internal Ultra3 SCSI disk drive backplane). VHDCI may require a mini 68-pin connector or FC 2118 mini 68-pin to 68-pin 0.3 meter cable as an additional feature.
- One Single Ended (SE) SCSI port on POWER4+ models.
- Three serial. Serial port 1 (S1) has two physical connectors, one RJ-48 in front and a 9-pin D-shell in the rear. The use of the front port disables the rear S1 port.
- One parallel (the parallel port is not accessible when in LPAR mode).
- Keyboard and mouse.

I/O expansion drawer

One or two optional 7311-D20 I/O drawers expand the Model 6C4 by providing additional PCI-X slots and disk drive bays. Keep in mind any possible LPAR requirements when configuring the disk backplane.

- Seven hot-plug PCI-X slots, 64-bit, 133 MHz, 3.3 volt
  - For more information on PCI-X, see Section 2.4, “PCI-X slots and adapters” on page 18.
- Up to 12 hot-swappable disk drive bays
  - The optional disk backplane (FC 6429) consists of two 6-pack enclosures. Each 6-pack must be connected to either an Ultra3 SCSI (FC 6203) or Ultra3 SCSI RAID adapter (FC 2498). A single SCSI adapter must reside in slot 7 or two SCSI adapters located in slot 4 and slot 7 of the I/O drawer and requires a SCSI cable for each 6-pack (FC 4257). For additional information, see “I/O” on page 30.
  - To support the I/O drawer, FC 9575 is required for new system builds and FC 6575 is required for existing systems (available through an MES upgrade). FC 6575 must be installed by a service representative.

Additional features

The following additional features are available:

- 3D graphics accelerators for the 6C4 and 6E4, and input devices for the Model 6E4.
- SP Switch2 adapter for the Model 6C4.
- A USB adapter, USB keyboards, and a USB mouse are available.
- Optional hot-plug redundant power supplies for the p630 and 7311-D20 (requires the addition of redundant cooling for the p630).
- Optional redundant hot-plug cooling fans for the p630 (required on POWER4+ systems).
- Integrated service processor.
- AIX 5L license included (except in pSeries 630 Linux ready Express Configurations).
- Support for 32-bit and 64-bit applications.

> **Note:** Before completing an order, you need to determine whether a diskette drive, mouse, national keyboard, graphics adapter, monitor, or additional features are required.

1.3 Model types

Figure 1-3 shows the package layout for the Models 6C4 and 6E4.

[4] Very High Density Cable Interconnect (VHDCI)

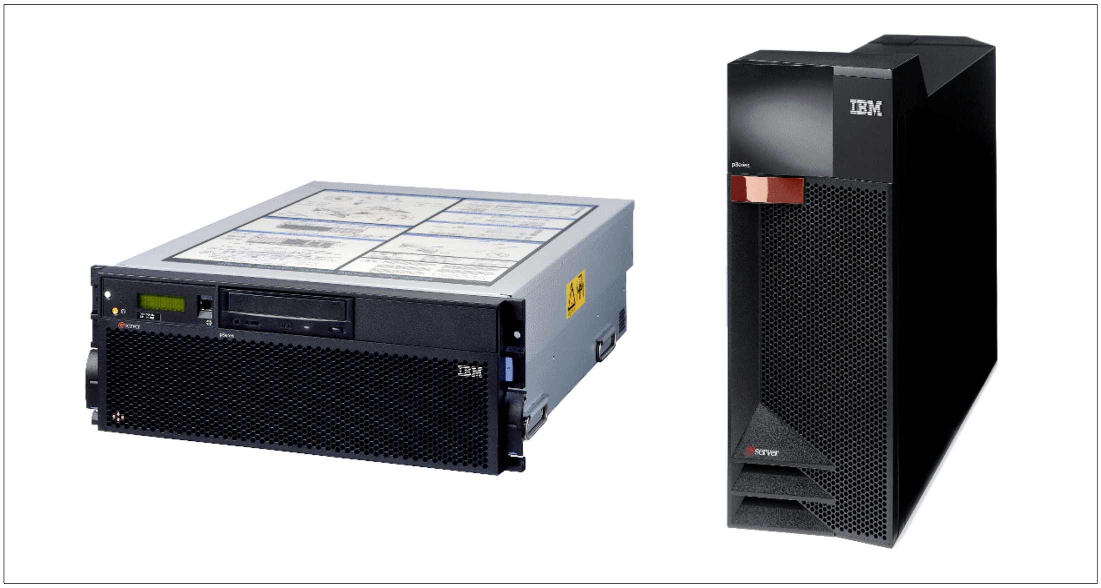


Figure 1-3   Models 6C4 (left) and 6E4 (right) physical packaging

1.3.1 Model 6C4 rack-mounted server

The Model 6C4 is a 4U rack-mounted server and is intended to be installed in a 19-inch rack, thereby enabling efficient use of computer room floor space. If the IBM 7014 T42 rack is used to mount the Model 6C4, then it is possible to place up to ten systems in an area of 644 mm (25.5 inches) x 1147 mm (45.2 inches).

Each system is delivered preconfigured as ordered. In the case of the pSeries 630 Linux ready Express Configurations, the OS is not installed. Included with the Model 6C4 rack-mounted server packaging will be all of the components and instructions necessary to enable installation in a 19-inch rack using suitable tools. This Model 6C4 is designated as a customer setup system and requires three persons (due to weight and safety issues) to be present to install the unit into the rack.

Each Model 6C4 is shipped with a template that helps each server be mounted into the desired position in the rack.

A service call is required for installation of system boards FC 6575 (RIO enabled planar), FC 6576 (LPAR enabled 4-slot riser), FC 6556 (6-slot riser card), FC 5132 or 5127 (the 2-way processor cards), and POWER4 to POWER4+ conversions; all other system MES upgrades can be performed by the customer.

The SP Switch2 Adapter (FC 8397) is available on this model.

On 1- or 2-way processor equipped 6C4 models, an optional POWER GXT4500P (FC 2842) or POWER GXT6500P (FC 2843) graphics accelerator provides advanced 3D graphics capabilities. A 2D graphics accelerator is available (FC 2848) and is supported in LPAR mode, however, SMS, firmware menus, and other functions that appear before AIX starts are only available from the HMC. See 2.4.4, “Graphics adapters” on page 19, for information on supported levels of adapters and placement guidelines.

1.3.2 Model 6E4 deskside server

The Model 6E4 doubles as a deskside server or workstation, ideal for environments requiring the user to have local access to the machine. A typical example of this would be applications requiring a native graphics display.

Each system will be delivered preconfigured as ordered. In the case of the pSeries 630 Linux ready Express Configurations, the OS is not installed. The system is designed to be set up by the customer and, in most cases, will not require the use of any tools. Full setup instructions are included with the system.

A service call is required for installation of system boards FC 6576 (LPAR enabled 4-slot riser), FC 6556 (6-slot riser card), FC 5132 or FC 5127 (the 2-way processor cards) and POWER4 to POWER4+ conversions; all system MES upgrades can be performed by the customer.

On 1- or 2-way processor equipped 6E4 models, an optional POWER GXT4500P (FC 2842) or POWER GXT6500P (FC 2843) graphics accelerator provides advanced 3D graphics capabilities, and the Spaceball (FC 8422) and Magellan XT (FC 8423) provide input and control of 3D applications. FC 2842 and FC 2843 accelerators are supported only in full system partition mode (non-LPAR). A 2D graphics accelerator is available (FC 2848) and is supported in LPAR mode; however, graphical SMS, firmware menus, and other functions that appear before AIX starts are only available from the HMC. See 2.4.4, “Graphics adapters” on page 19, for information on supported levels of adapters and placement guidelines.

1.4 System racks

The following description provides an overview of racks available from IBM in which the Model 6C4 can be mounted. There is no feature available to convert a deskside/desktop model to a rack-mounted model.

The Enterprise Rack Models T00 and T42 are 19-inch wide racks for general use with pSeries and RS/6000 rack-based or rack drawer-based systems. The rack provides increased capacity, greater flexibility, and improved floor space utilization.

If a pSeries or RS/6000 system is to be installed in a non-IBM rack or cabinet, you should ensure that the rack to be used conforms to the EIA standard EIA-310-D.

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

1.4.1 IBM RS/6000 7014 Model T00 Enterprise Rack

The 1.8 meter (71 inches) Model T00 is compatible with past and present pSeries and RS/6000 racks, and is designed for use in all situations that have previously used the older rack Models R00 and S00. The T00 rack has the following features:

- 36 EIA units (36U) of usable space.
- Optional removable side panels.
- Optional highly perforated front door.
- Optional side-to-side mounting hardware for joining multiple racks.
- Increased power distribution and weight capacity.

- Standard black or optional white color in OEM format.
- Optional reinforced (ruggedized) rack feature provides added earthquake protection with modular rear brace, concrete floor bolt-down hardware, and bolt-in steel front filler panels.
- Optional rack status beacon (FC 4690). This beacon is designed to be placed on top of a rack and cabled to servers such as p630 and other components such as I/O drawer 7311-D20 inside the rack. Servers can be programmed to illuminate the beacon in response to any detected problems or changes in system status.

  To use this beacon, a rack status beacon junction box (FC 4693) should be selected to connect multiple servers and I/O drawers to the beacon. This feature provides six input connectors and one output connector for the rack. To connect the servers or other components to the junction box or the junction box to the rack, status beacon cables (FC 4691) are necessary. Multiple junction boxes can be linked together in a series using daisy chain cables (FC 4692).
- Support of both AC and DC configurations.
- Up to six Power Distribution Units (PDUs). See 1.4.3, “AC Power Distribution Units for rack models T00 and T42” on page 9.
- DC rack height is increased to 1926 mm (75.8 inches) due to the presence of the power distribution panel fixed to the top of the rack.
- Weight:
  - T00 base empty rack: 244 kg (535 pounds)
  - T00 full rack: 816 kg (1795 pounds)

1.4.2 IBM RS/6000 7014 Model T42 Enterprise Rack

The 2.0 meter (79.3 inches) Model T42 is the rack that will address the special requirements of customers who want a tall enclosure to house the maximum amount of equipment in the smallest possible floor space. The features that differ in the Model T42 rack from the Model T00 include the following:

- 42 EIA units (42U) of usable space.
- AC power support only.
- Weight:
  - T42 base empty rack: 261 kg (575 pounds)
  - T42 full rack: 930 kg (2045 pounds)

1.4.3 AC Power Distribution Units for rack models T00 and T42

For rack Models T00 and T42, different Power Distribution Units (PDUs) are available. Previously, the only AC PDUs for these racks were PDUs with six outlets (FC 9171, 9173, 9174, 6171, 6173, and 6174). Four PDUs can be mounted vertically and two additional PDUs horizontally on the bottom of the rack requiring 1U of rack space for each horizontally mounted PDU. Each PDU requires a separate AC supply. These PDUs provide a maximum of 36 power connections.

In addition to the six outlet PDUs, new PDUs with nine outlets are available. A T42 rack configured for the maximum number of power outlets would have six PDUs (two mounted horizontally requiring 2U of rack space), for a total of 54 power outlets. T00 racks do not allow horizontal placement of these PDUs and are therefore limited to a total of four nine-outlet PDUs or 36 outlets.

Figure 1-4 shows the new PDU.

For the Model 6C4, the initial PDU could be selected out of one of the following when using the new PDUs:

- FC 9176 Power Distribution Unit, base/side mount, single phase, L6-30 connector
- FC 9177 Power Distribution Unit, base/side mount, single phase, IEC-309 connector
- FC 9178 Power Distribution Unit, base/side mount, three phase, IEC-309 connector

Additional PDUs with nine power outlets can be added to a configuration out of the following, based on which initial PDU (FC 9176, 9177, or 9178) was selected:

- FC 7176 Power Distribution Unit, side mount, single phase, L6-30 connector
- FC 7177 Power Distribution Unit, side mount, single phase, IEC-309 connector
- FC 7178 Power Distribution Unit, side mount, three phase, IEC-309 connector

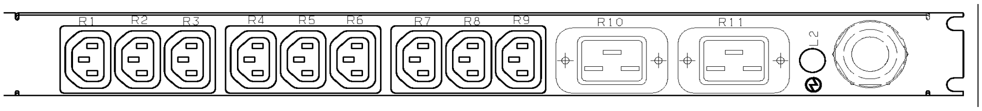


Figure 1-4 New PDU for p630 and other pSeries servers

1.4.4 OEM racks

The Model 6C4 can be installed in a suitable OEM rack, provided that the rack conforms to the EIA-310-D standard. This standard is published by the Electrical Industries Alliance, and a summary of this standard is available in the publication *Site and Hardware Planning Information*, SA38-0508. An online copy of this document can be found at:

http://www.ibm.com/servers/eserver/pseries/library/hardware_docs

Key points mentioned in this standard are as follows:

- Any rack used must be capable of supporting 15.9 kg (35 pounds) per EIA unit (44.5 mm (1.75 inch) of rack height).
- To ensure proper rail alignment, the rack must have mounting flanges that are at least 494 mm (19.45 inches) across the width of the rack and 719 mm (28.3 inches) between the front and rear rack flanges.
- It may be necessary to supply additional hardware, such as fasteners, for use in some manufacturers' racks.

1.4.5 Rack-mounting rules

There are rules that should be followed when mounting the Model 6C4 or the I/O drawer (7311-D20) into a rack. The primary rules are as follows.

Model 6C4

The rules for the Model 6C4 are:

- The Model 6C4 is designed to be placed at any location in the rack. For rack stability reasons, it is advisable to start filling a rack from the bottom.
- Any remaining space in the rack can be used to install other systems or peripherals provided that the maximum permissible weight of the rack is not exceeded.

- Before placing a Model 6C4 into the service position, it is essential that the rack manufacturer's safety instructions have been followed regarding rack stability.
- A Model 6C4 is 4U in height, so a maximum of nine Model 6C4s fits in a T00 rack, or ten Model 6C4s in a T42 rack.

7311-D20 I/O drawer

The rules for the 7311-D20 I/O drawer are:

- A maximum of nine 4U Model D20s can be mounted in a T00 rack or a maximum of ten Model D20s in a T42 rack.
- The 7311-D20 I/O drawer is designed to be placed at any location in the rack. For rack stability reasons, it is advisable to start filling an empty rack from the bottom and place I/O drawers above system units.
- The I/O drawers could be in the same rack as the Model 6C4 server or in an adjacent rack, although it is recommended that the I/O drawers be located in the same rack as the server for service considerations.
- The I/O drawer is an IBM service representative installable item.

1.4.6 Flat panel display options

For rack-mounted systems, the IBM 7316-TF2 Flat Panel Console Kit may be installed in the system rack. This 1U (EIA) console uses a 15-inch thin film transistor (TFT) LCD with a viewable area of 304.1 mm x 228.1 mm and a 1024 x 768 resolution.

Note: It is recommended that the 7316-TF2 be installed between EIA 20 to 25 of the rack for ease of use. The 7316-TF2 or any other graphics monitor requires a GXT135P Graphics Adapter (FC 2848) to be installed in the server, or other graphics adapter, if supported.

The 7316-TF2 Flat Panel Console Kit has the following attributes:

- Flat panel color monitor.
- Rack tray for keyboard, monitor, and optional VGA switch with mounting brackets.
- IBM Space Saver 2 14.5-inch Keyboard that mounts in the Rack Keyboard Tray and is available as a feature in sixteen language configurations (the track point mouse is integral to the keyboard).

The L200P Flat-Panel Monitor (FC 3636) provides a desktop flat panel display option. The L200P is a 20.1-inch TFT LCD digital screen and a maximum resolution of 1600 x 1200 pels at 75 Hz in analog mode and 60 Hz in digital mode.

1.4.7 VGA switch

The VGA switch for the IBM 7316-TF2 allows for the connection of up to eight servers to a single keyboard, display, and mouse.

To help minimize cable clutter, multi-connector cables in lengths of 7, 12, and 20 feet are available. These cables can be used to connect the graphics adapter (required in each attached system), keyboard port, and mouse port of the attached servers to the switch, or to connect between multiple switches in a tiered configuration. Using a two-level cascade arrangement, as many as 64 systems can be controlled from a single point.

This dual-user switch allows attachment of one or two consoles, one of which must be an IBM 7316-TF2. An easy-to-use graphical interface allows fast switching between systems and supports six languages (English, French, Spanish, German, Italian, or Brazilian Portuguese).

The VGA switch is only 1U high and can be mounted in the same tray as the IBM 7316-TF2, thus conserving valuable rack space. It supports a maximum video resolution of 1600 x 1280, which facilitates the use of graphics-intensive applications and large monitors.

1.5 Statements of direction

All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only. The following are IBM’s statements of direction for the pSeries 630.

1.5.1 NEBS and -48 volts

IBM is planning to provide NEBS[5] Level 3 compliance for the pSeries 630 Model 6C4 and -48V DC power in the first half of 2003.

1.5.2 PSSP support provided on AIX 5L Version 5.2

To help you plan your Cluster 1600 migrations to AIX 5L Version 5.2, the following information is being made available. IBM intends to provide support for AIX 5L Version 5.2 with the PSSP for AIX product on Cluster 1600 systems in 2003.

[5] Network Equipment Building System

2

Architecture and technical overview

This chapter discusses the overall system architecture represented by Figure 2-1. The major components of this diagram will be described in the following sections. The bandwidths provided throughout this section are theoretical maximums provided for reference. It is always recommended to obtain real-world performance measurements using production workloads.

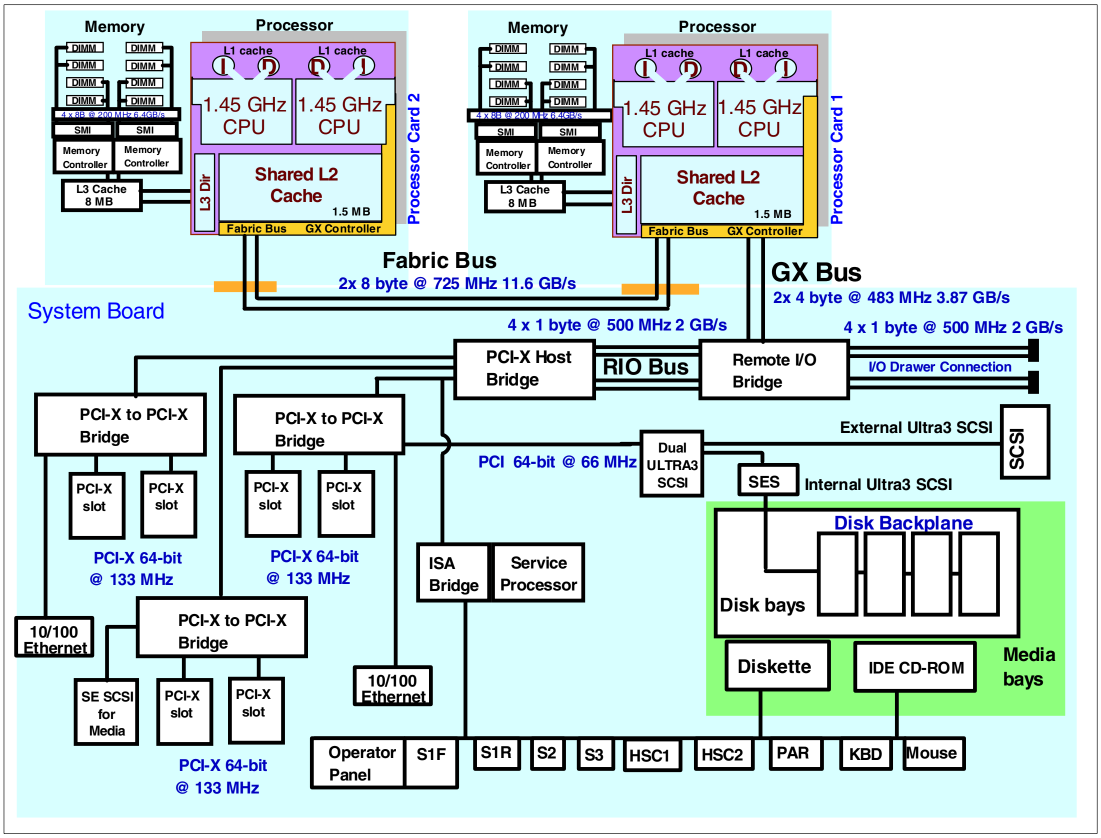


Figure 2-1 Conceptual diagram of the Models 6C4 and 6E4 POWER4+ system architecture

2.1 Processor and cache

The Models 6C4 and 6E4 processor subsystems consist of POWER4 processor(s) running at 1.0 GHz or POWER4+ processor(s) running at 1.2 GHz or 1.45 GHz. These are related to the processors used in the IBM @server pSeries 690 and 650. However, unlike the Multichip Module (MCM) packaging used in the Model 690, these systems use a Single Chip Module (SCM) containing either one or two processor cores (CPUs), with each SCM permanently mounted on a processor card. Models 6C4 and 6E4 can contain one or two processor cards, giving the option of one, two, or four CPUs per system.

One key difference is that the chip-to-chip fabric bus (which was used between chips on the same MCM in the Model 690) is no longer relevant in the SCM. It is replaced by a module-to-module fabric. A logical diagram of the pSeries 630 SCM and the pSeries 690 MCM is shown in Figure 2-2 for comparison.

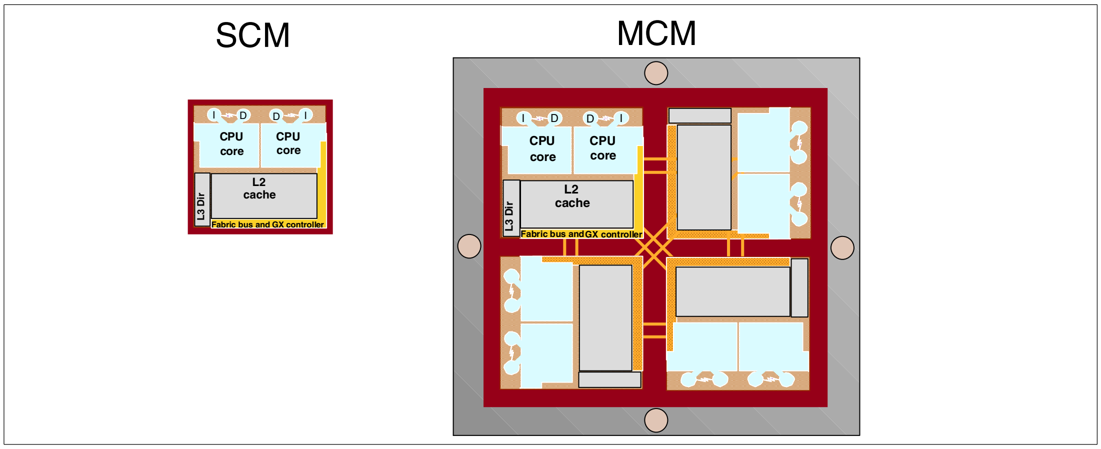


Figure 2-2 Comparison between SCM and MCM

Each SCM is a Ceramic Column Grid Array (CCGA) package where the chip carrier is raised slightly from its board mounting by small metal solder columns that provide the required connections and improved thermal resilience characteristics. As well as the SCM, each processor card also contains the L3 cache and the memory DIMMs, as shown in Figure 2-3. The processor card is mounted in a rugged metal enclosure (*book*) that protects and secures the card (both in and out of the server), and helps manage airflow used for cooling.

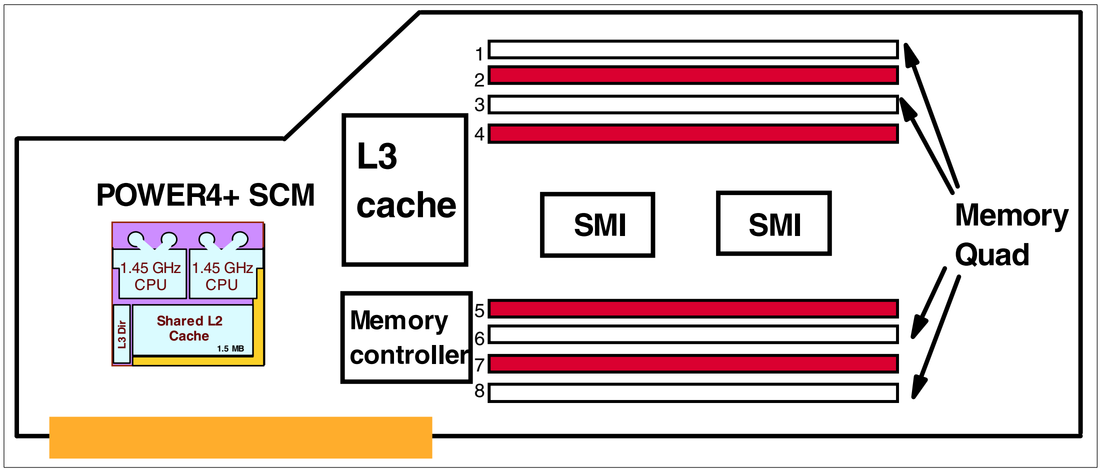


Figure 2-3 POWER4+ Processor card layout

Memory access is through the on-chip Level 2 cache (L2) and Level 3 (L3) cache directory controller to the off-chip L3 cache and finally through the memory controller and synchronous memory interface (SMI) to the memory DIMMs, as represented in Figure 2-4.

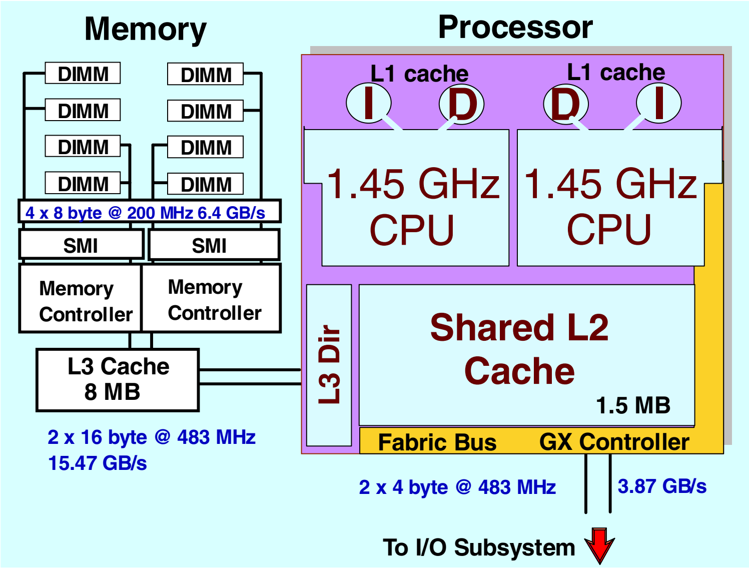


Figure 2-4 Conceptual diagram of POWER4+ processor and memory subsystem

2.1.1 L1, L2, and L3 cache

The POWER4+ storage subsystem consists of three levels of cache and the memory subsystem. The first two levels of cache are onboard the POWER4+ chip. The first level is 64 KB of Instruction (I) and 32 KB of Data (D) cache per processor core. The second level is 1.5 MB of L2 cache on the POWER4+ and 1.44 MB on the POWER4 chip. All caches have either full ECC[1] or parity protection on the data arrays, and the L1 cache has the ability to re-fetch data from the L2 cache in the event of soft errors detected by parity checking.

A 2-way configuration using two 1-way processor cards offers better performance than a configuration using a single 2-way processor card because the maximum capacity of memory is doubled from 16 GB to 32 GB and the Level 2 (L2) and Level 3 (L3) cache on each card is dedicated to a single processor. A 2-way configuration using 2-way processor card shares the L2 and L3 caches. A comparative table of 2-way configurations is provided in Table 2-1.

Table 2-1 2-way configurations comparative table

2-way configuration	Rperf[a]	Memory GB	L2 Cache MB	L3 Cache MB
One 2-way 1.0 GHz POWER4 processor card	3.29	16	1.44 shared	32 shared
Two 1-way 1.0 GHz POWER4 processor cards	3.69	32	1.44 per processor	32 per processor
One 2-way 1.2 GHz POWER4+ processor card	3.73	16	1.5 shared	8 shared

[a] Source: IBM @serverpSeries and IBM RS/6000 Facts and Features - March 27, 2003

[1] ECC single error correct, double error detect

2-way configuration	Rperf[a]	Memory GB	L2 Cache MB	L3 Cache MB
Two 1-way 1.2 GHz POWER4+ processor cards	4.18	32	1.5 per processor	8 per processor
One 2-way 1.45 GHz POWER4+ processor card	4.41	16	1.5 shared	8 shared
Two 1-way 1.45 GHz POWER4+ processor cards	4.94	32	1.5 per processor	8 per processor

[a] Source: IBM @serverpSeries and IBM RS/6000 Facts and Features - March 27, 2003

Many applications, such as high performance computing (HPC) and commercial and professional workstation environments, may take advantage of the larger memory capacity and increased L2 and L3 cache per processor available with the two 1-way processor configurations to achieve improved performance. The compute intensive or I/O intensive operations that are typical of these kinds of applications often benefit from large memory capacity and more L2/L3 cache per processor.

The Level 3 (L3) cache consists of two components: The L3 cache controller/directory and the L3 data array. The L3 cache controller/directory is on the POWER4 and POWER4+ chip, and the L3 data array, which consists of 32 MB (POWER4) or 8 MB (POWER4+) of embedded DRAM (eDRAM), is located on a separate module mounted on the processor card.

2.1.2 PowerPC architecture

The Models 6C4 and 6E4 systems comply with the RS/6000 platform architecture, which is an evolution of the PowerPC Common Hardware Reference Platform (CHRP) specifications.

2.1.3 Copper and CMOS technology

The POWER4+ processor chip takes advantage of IBM's leadership technology. It is made using IBM 0.13-µm-lithography CMOS[2] fabrication with seven levels of copper interconnect wiring. POWER4+ also uses Silicon-on-Insulator (SOI) technology to allow a higher operating frequency for improved performance, yet with reduced power consumption and improved reliability compared to processors not using this technology.

2.1.4 Processor clock rate

The Models 6C4 and 6E4 operate with a processor clock rate of 1.0 GHz for POWER4 systems and 1.2 GHz or 1.45 GHz for POWER4+ systems.

To determine the processor characteristics on a running system, use one of the following commands:

`lsattr -El proc`*`X`* Where *X* is the number of the processor, for example, proc0 is the first processor in the system. The output from the command[3] would be similar to the following (False, as used in this output, signifies that the value cannot be changed through an AIX command interface):

[2] Complementary Metal Oxide Semiconductor

[3] The output of the `lsattr` command has been expanded with AIX 5L to include the processor clock rate.

```
state enable          Processor state  False
type powerPC_POWER4   Processor type   False
frequency 100000000   Processor Speed  False
```

`pmcycles -m` This command (AIX 5L Version 5.1 and higher) uses the performance monitor cycle counter and the processor real-time clock to measure the actual processor clock speed in MHz. Here is the output of a 2-way p630 1.00 GHz system:

```
Cpu 0 runs at 1000 MHz
Cpu 1 runs at 1000 MHz
```

Note: The **`pmcycles`** command is part of the bos.pmapi fileset. First check if that component is installed using the **`lslpp -l bos.pmapi`** command.

2.2 Memory

The conceptual diagram of the memory subsystem of the Models 6C4 and 6E4 is shown in Figure 2-4 on page 15. As shown, there are four 8-byte data paths from the memory controller to the memory with an aggregated bandwidth of 6.4 GB/s on each processor card. Each processor card can hold up to eight double data rate (DDR) synchronous DRAM (SDRAM) DIMM memory cards, which must be populated in quads.

DDR memory can theoretically double memory throughput at a given clock speed by providing output on both the rising- and falling-edges of the clock signal (rather than just on the rising edge).

Memory must be balanced across the two-processor cards for best performance, for example, a 1 GB 4-way configuration is not recommended.

The p630 supports memory affinity (requires the 12/06/2002 driver), which is the ability to reserve memory on a processor card to be used exclusively for the processors on that card, as opposed to crossing the module-to-module fabric bus and accessing the memory on another card (4-way systems only). The default setting is off.

2.2.1 Memory options

The following memory features for the Models 6C4 and 6E4 were available at the time this publication was written:

FC 4451	1024 MB (4 x 256 MB) 208 pin 8 ns DDR SDRAM DIMMs
FC 4452	2048 MB (4 x 512 MB) 208 pin 8 ns DDR SDRAM DIMMs
FC 4453	4096 MB (4 x 1024 MB) 208 pin 8 ns stacked DDR SDRAM DIMMs
FC 4454	8192 MB (4 x 2048 MB) 208 pin 8 ns stacked DDR SDRAM DIMMs

Each memory feature consists of four DIMMs, or quad. The memory installed on processor card 1 should match the size of the memory installed on processor card 2.

2.3 System buses

The system bus from the processors to the memory subsystem is 2 x 16 bytes at 1/3 CPU clock speed (or 483.3 MHz on 1.45 GHz POWER4+ systems) for an aggregate data rate of 15.47 GB/s. For a system with two processor cards, the fabric bus connects between

POWER4 chips on each card. All traffic to and from the I/O subsystem is through the GX bus on the first processor card. All system, I/O, and PCI-X buses support parity error detection.

2.3.1 GX bus and fabric bus

The GX controller (embedded in the POWER4+ chip) is responsible for controlling the flow of data through the GX bus. The GX bus is a high-frequency, single-ended, unidirectional, point-to-point bus. Both data and address information are multiplexed onto the bus, and for each path there is an identical bus for the return path. The GX bus has dual 4-byte paths.

The first processor card connects to the GX bus through its GX controller. The GX bus is connected to a Remote I/O (RIO) bus on the system board through a RIO bridge chip. Each RIO bus provides 1 byte at 500 MHz in each direction, or 1 GB/s bidirectionally.

The second processor card has access to the GX bus using a module-to-module fabric bus, which connects it to the first processor card. The fabric bus is similar in nature to the GX bus, but has dual 8-byte paths.

Characteristics of the GX bus and fabric bus are provided in Table 2-2.

Table 2-2 GX bus and fabric bus characteristics

Processor	1.0 GHz POWER4	1.2 GHz POWER4+	1.45 GHz POWER4+
GX bus frequency	333 MHz	400 MHz	483.3 MHz
GX bus data rate	2.67 GB/s	3.2 GB/s	3.87 GB/s
Fabric bus frequency	500 MHz	600 MHz	725 MHz
Fabric bus data rate	8 GB/s	9.6 GB/s	11.6 GB/s

2.3.2 PCI host bridge and PCI bus

The remote I/O bridge contains a series of 1-byte busses grouped in pairs called ports. One of the busses in the pair is for inbound data transfers (500 MB/s), while the other bus is for outbound data transfers (500 MB/s). There is a total of four ports. Two ports are for the internal I/O and two ports are used to attach to any I/O drawers in a loop using RIO cables. The two internal ports connect to a PCI-X Host Bridge (PHB). The PHB chip acts as a bridge between the RIO bus and two PCI-X to PCI-X bridges, which fan out to integrated I/O controllers and slots.

Integrated devices include two Ethernet controllers on the system board and a dual Ultra3 SCSI controller located on the PCI-slot riser card. These integrated devices and I/O slots are further described in the following section.

2.4 PCI-X slots and adapters

PCI-X is the latest version of PCI bus technology, using a higher clock speed (133 MHz) to deliver a bandwidth of up to 1 GB/s. The PCI-X slots in the Model 6C4 and 6E4 systems support hot-plug and Extended Error Handling (EEH). EEH-capable adapters respond to a specially generated data packet from a PCI slot with a problem. This packet is analyzed by the system firmware, which then allows the device driver to reset the adapter or slot, isolating the error and reducing the need for a system reboot.

PCI-X slots also support existing 3.3 volt PCI adapters. For a full list of the adapters that are supported on the Model 6C4 and 6E4 systems, and for important information regarding adapter placement, see the *RS/6000 and pSeries PCI Adapter Placement Reference*, SA38-0538, for additional information. You can find this publication at:

http://www.ibm.com/servers/eserver/pseries/library/hardware_docs/pci_adp_pl.html

2.4.1 64-bit and 32-bit adapters

Since PCI-X slots support existing 3.3 volt PCI adapters, that enables them to support both 64-bit and 32-bit adapters.

Choosing between 32-bit and 64-bit adapters influences slot placement and affects performance. Higher-speed adapters use 64-bit slots because they can transfer 64 bits of data for each data transfer phase. 32-bit adapters can typically function in 64-bit PCI slots; however, 32-bit adapters still operate in 32-bit mode and achieve no performance advantage in a 64-bit slot.

Note: Models 6C4 and 6E4, as well as the 7311-D20 I/O drawer, do not contain any 5 volt slots. Therefore, adapters limited exclusively to 5 volt signaling must not be used.

2.4.2 Internal Ultra3 SCSI controllers

The integrated Ultra3 SCSI controllers provide one external Ultra3 SCSI VHDCI port and one internal SCSI to support the disk drive backplane and associated internal disk subsystem. Ultra3 SCSI supports single-ended or low-voltage differential (LVD) devices at up to 160 MB/s[4].

A SCSI Enclosure Services (SES) device supports the disk bay hot-swappable features. The SES processor is the SCSI hot-swap manager; it provides the control mechanism for the device hot-swap options such as identify/replace/remove.

POWER4+ systems with the 6-slot riser also have an integrated Single Ended (SE) SCSI controller to handle SCSI devices installed in the media bays. For more information, see “I/O” on page 30.

2.4.3 LAN adapters

Since the Models 6C4 and 6E4 function as servers, they are usually connected to a local area network (LAN). The two internal 10/100 Mbps Ethernet integrated controllers, situated on the system board, can be used to accomplish that.

Tip: In conjunction with certain network switches, you can use the Cisco Systems' EtherChannel feature of AIX to build up one virtual Ethernet interface with increased bandwidth using two to four Ethernet interfaces (adapters or integrated).

LAN adapters include: Gigabit and 4-port Ethernet, token-ring, and ATM. IBM supports an installation with NIM using Ethernet or token-ring adapters (use chrp as the platform type).

2.4.4 Graphics adapters

The p630 supports the POWER GXT135P (FC 2848) 2D graphics adapter, and the POWER GXT4500P (FC 2842) or POWER GXT6500P (FC 2843) 3D graphics accelerators. The

[4] SCSI T10 Technical Committee, http://www.t10.org

Model 6E4 also supports the Spaceball (FC 8422) and Magellan XT (FC 8423) input devices. The following are the guidelines:

- A maximum of one FC 2842 or FC 2843 is supported per Model 6E4.
- FC 2842 and FC 2843 are not supported in LPAR mode.
- The blue handle must be removed on FC 2842 and FC 2843 for use.
- FC 2842 or FC 2843 may not be placed in slots 1 or 2 or combined with the SP Switch2 adapter FC 8397 in the same machine.
- FC 2842 or FC 2843 must be placed in slot 3 or 4.
- POWER4 systems (FC 5131 or FC 5132) have following slot restrictions:
  - If FC 2842 or FC 2843 is placed in slot 3, slot 4 must remain empty.
  - If FC 2842 or FC 2843 is placed in slot 4, slot 3 must remain empty.
  - FC 2842 or FC 2843 is not allowed in slots 1 and 2.
- POWER4+ systems (FC 5126, FC 5127, FC 5133 or FC 5134) have following slot restriction:
  - FC 2842 or FC 2843 is not allowed in slots 1, 2, 5, or 6.
- Only a single FC 2848 is allowed when a FC 2842 or FC 2843 is installed.
- A maximum of four FC 2848s are supported on the Model 6C4 and Model 6E4.

FC 2842 must be at level 00P4476 and FC 2843 must be at level 00P4473 in order to function correctly in the Model 6E4.

2.5 Internal storage

The Models 6C4 and 6E4 have two internal media bays. These bays can be populated by the devices listed in Table 2-3.

Table 2-3 List of internal media options

Feature Code	Description	Comments
2633	650 MB IDE 48x CD-ROM Drive	This drive is attached to an IDE cable that is part of the base system.
2634	DVD-ROM Drive	This drive is attached to an IDE cable that is part of the base system. DVD-ROM operating system support requires AIX 5L Version 5.2.
2605	Diskette Drive	This diskette drive, when fitted, occupies a full media bay.
2623	DVD-RAM Drive 4.7 GB capacity	Systems with the 6-slot riser will use the integrated SE SCSI adapter and FC 4250 for connection. Systems with 4-slot riser cards will require FC 6203 with 4260. This SCSI card will necessitate the use of a PCI slot within the machine.
6158	4 mm 20/40 GB Tape Drive	Systems with the 6-slot riser will use the integrated SE SCSI adapter and FC 4250 for connection. Systems with 4-slot riser cards will require FC 6203 with 4260. This SCSI card will necessitate the use of a PCI slot within the machine.
Note: A single SCSI adapter, FC 6203 with FC 4260 (2-drop connector cable) or the integrated adapter on 6-slot machines, is sufficient for supporting SCSI devices in both the internal media bays.		

Feature Code	Description	Comments
6134	8 mm 60/150 GB Tape Drive	Systems with the 6-slot riser will use the integrated SE SCSI adapter and FC 4250 for connection. Systems with 4-slot riser cards will require FC 6203 with 4260. This SCSI card will necessitate the use of a PCI slot within the machine.
6120	80/160 GB VXA Tape Drive	Systems with the 6-slot riser will use the integrated SE SCSI adapter and FC 4250 for connection. Systems with 4-slot riser cards will require FC 6203 with 4260. This SCSI card will necessitate the use of a PCI slot within the machine.
Note: A single SCSI adapter, FC 6203 with FC 4260 (2-drop connector cable) or the integrated adapter on 6-slot machines, is sufficient for supporting SCSI devices in both the internal media bays.		

2.5.1 Hot-swappable SCSI disks

The Models 6C4 and 6E4 can have up to four hot-swappable drives in the front hot-swappable disk bay. The hot-swap process is controlled by the SCSI enclosure service, ses0, which is located on the PCI riser card rather than on the SCSI backplane. The PCI riser card has an integrated Ultra3 SCSI adapter. This adapter has two separate controllers: One feeding the hot-swappable bay and the other using an internal cable to supply the external VHDCI mini 68-pin SCSI connector. The hot-swappable bays can use the devices listed in Table 2-4.

Table 2-4 Hot-swappable disk options

Feature Code	Description
3157	18.2 GB 10,000 RPM Ultra3 SCSI hot-swappable disk drive
3158	36.4 GB 10,000 RPM Ultra3 SCSI hot-swappable disk drive
3277	36.4 GB 15,000 RPM Ultra3 SCSI hot-swappable disk drive
3159	73.4 GB 10,000 RPM Ultra3 SCSI hot-swappable disk drive
3278	73.4 GB 15,000 RPM Ultra3 SCSI hot-swappable disk drive
3275	146.8 GB 10.000 RPM Ultra3 SCSI hot-swappable disk drive

Disk drive fault tracking of transient errors can alert the system administrator of an impending disk failure before it impacts system operation.

2.5.2 Other hot-plug options

The Models 6C4 and 6E4 also give you the ability to concurrently change or add PCI adapters and the disks within the system.

The addition, removal, or changing of a PCI adapter can be accomplished by using a system management tool such as Manage PCI Hot-Plug Slots (using the Web-based System Manager) or PCI Hot-Plug Manager (using SMIT). The PCI hot-plug tasks can be accomplished also by using the Hot-Plug Task of the online diagnostics task selection menu (using the **diag** command). Each of these tools provide a method by which a PCI slot can be identified first, powered off to enable removal or insertion of an adapter, and then powered back on to enable the device to be configured.

Prior to the hot-swap of a disk in the hot-swappable bay, all necessary operating system actions must be undertaken to ensure that the disk is capable of being deconfigured.

Once the disk drive has been deconfigured, the SCSI enclosure device will power off the slot, enabling safe removal of the disk. You should ensure that the appropriate planning has been given to any operating system related disk layout, such as the AIX Logical Volume Manager, when using disk hot-swap capabilities. For more information, see *Problem Solving and Troubleshooting in AIX 5L*, SG24-5496.

2.5.3 Boot options

Both Models 6C4 and 6E4 handle the boot process in a way that is similar to other pSeries servers.

The initial stage of the boot process is to establish that the machine has powered up correctly and the memory and CPUs are functioning correctly. This sequence of events is performed by the service processor of the Models 6C4 and 6E4. Once the machine reaches the System Management Services (SMS) menus, all the necessary tests have been performed and the machine is scanning the bus for a boot source.

At this point, there are a number of possibilities:

CD-ROM, DVD-ROM, DVD-RAM These devices can be used to boot the system so that a system can be loaded, system maintenance performed, or stand-alone diagnostics performed.

Internal or external tape drives The media bay tape drive or any externally attached tape drive can be used to boot the system using **`mksysb`**, for example.

SCSI disk The more common method of booting the system is to use a disk situated in one of the hot-swap bays in the front of the machine. However, any external non-RAID SCSI attached disk could be used if required.

SSA disk The Models 6C4 and 6E4 support booting from an SSA disk either as an AIX system disk or as a RAID LUN. FC 6230 serial RAID adapters can be used to provide the boot support from a RAID-configured disk provided that the firmware level of the adapter is 7000 and higher. For more information on the SSA boot, see the SSA frequently asked questions located on the Web[5].

Note: Fastwrite must not be enabled on the boot resource SSA adapter.

SAN boot It is possible to boot these systems from a SAN using a 2 GB Fibre Channel Adapter (FC 6228) with microcode 3.22A1 or later installed on the adapter[6]. The IBM 2105 Enterprise Storage Server (ESS) is an example of a SAN-attached device that can provide a boot medium.

LAN boot Network boot and NIM installs can be used if required. An example of this would be in the ISP environment, where a high density of machines are installed in racks and it is impractical to use CD media.

[5] http://www.storage.ibm.com/hardsoft/products/ssa/faq.html#microcode

[6] http://techsupport.services.ibm.com/server/mdownload/download.html

2.6 I/O drawer

The POWER4+ Model 6C4 has six internal PCI-X slots and four disk bays in one 4-pack, which is sufficient for basic operation. If more PCI-X slots and disks are needed, especially well-suited for LPAR mode, up to two 7311-D20 I/O drawers can be added to the Model 6C4. For a Model 6E4, the attachment of the 7311-D20 I/O drawer is not supported.

> **Note:** Existing machines, which do not have 6 PCI-X adapter slots, will require an MES upgrade of the POWER4 Model 6C4 system planar in order to support the attachment of I/O drawers (FC 6575). This upgrade must be done by an IBM service representative. New orders use FC 9575 to indicate RIO and LPAR capability.

The 7311-D20 is a 4U full-size drawer, which must be mounted in a rack. It features seven hot-pluggable PCI-X slots and up to 12 hot-swappable disks arranged in two 6-packs. Redundant concurrently maintainable power and cooling is an optional feature (FC 6268).

The hot-plug mechanism is the same as on the Models 6F1, 6H1, or 6M1; therefore, PCI cards are inserted from the top of the I/O drawer down into the slot. The installed adapters are protected by plastic separators, designed to prevent grounding and damage when adding or removing adapters. Figure 2-5 shows an inside view of the 7311-D20 I/O drawer.

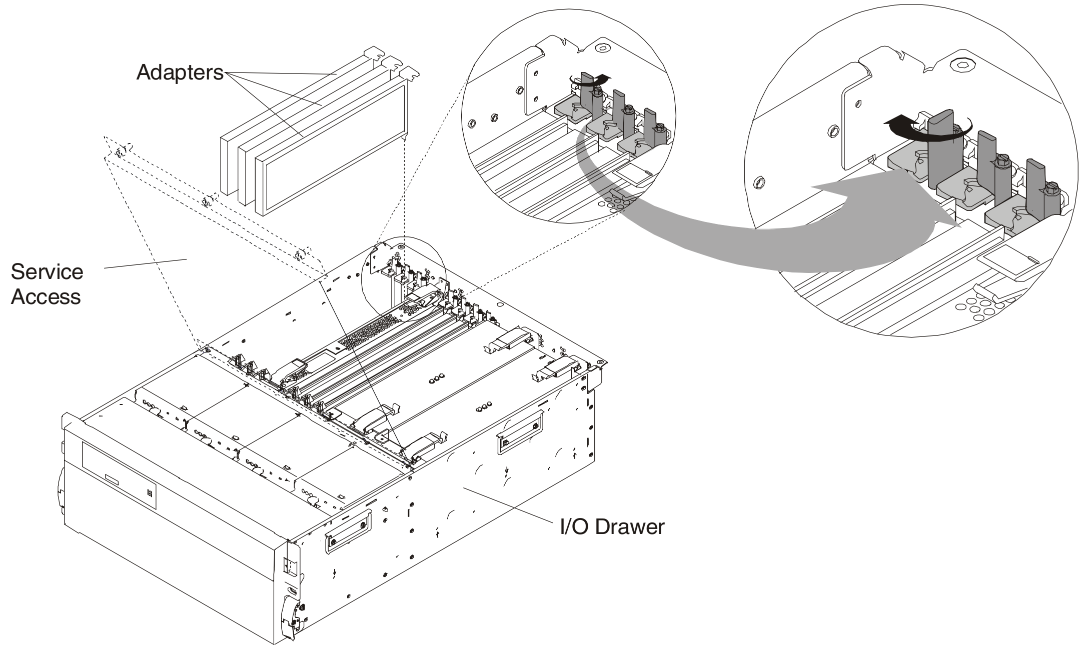


Figure 2-5 Inside view of 7311-D20 I/O drawer

2.6.1 I/O drawer ID assignment

A basic understanding of I/O drawer ID assignment is vital for locating devices and providing your service representative with proper information.

Each I/O drawer belonging to the Model 6C4 is uniquely identified by an ID, such as U0.*x*, where *x* is the I/O drawer ID starting from the number 2, since the Model 6C4 is identified as U0.1.

The first ID assignment calls all the I/O drawers attached to the Model 6C4 according to the sequence they have in the SPCN loop. At this time, the system firmware creates a table into NVRAM reporting the I/O drawer ID assignment against the I/O drawer system serial number.

If you have the system A with two I/O drawers, the assigned addresses are U0.2, and U0.3. If you want to move the first I/O drawer, U0.2, away for any reason, the system maintains the name of the other I/O drawer, U0.3, and its related parts, without any change of the AIX physical locations related to the adapters inside this drawer. If you add a brand new I/O drawer or an I/O drawer from system B that was never used with system A, system A assigns to the I/O drawer the first address after the last configured I/O drawer, in our case U0.4 and not U0.2. If you move back the removed I/O drawer U0.2, system searches a table to determine if the serial number is known. Since it is, the system recognizes it as U0.2 again.

2.6.2 I/O drawer PCI-X subsystem (7311-D20)

The Model 6C4 provides two RIO ports for connecting up to two I/O drawers in a single loop. Each RIO port operates at 500 MHz in bidirectional mode and is capable of passing up to eight bits of data in each direction on each cycle of the RIO port. If only one I/O drawer is connected to Model 6C4, the RIO loop runs in double barrel mode, meaning both cables are used by this I/O drawer, giving a total bandwidth of 2 GB/s for this drawer. One RIO port (capable of 1 GB/s) will be assigned to a PHB (slots 1-4) and the other RIO port (capable of 1 GB/s) will be assigned to the other PHB (slots 5-7). If two I/O drawers are connected to a Model 6C4, in normal operation each I/O drawer uses one cable, giving a total bandwidth of 1 GB/s for each I/O drawer to be shared between the internal devices in that drawer.

The PCI-X host bridge inside the I/O drawer provides two primary 64-bit PCI-X buses running at 133 MHz. Therefore, a maximum bandwidth of 1 GB/s is provided by each of the buses. To avoid overloading an I/O drawer, the *RS/6000 and pSeries PCI Adapter Placement Reference Guide*, SA38-0538, must be followed.

Each primary PCI-X bus is connected to a PCI-X-to-PCI-X bridge. The first PCI-X-to-PCI-X bridge provides four slots and the second PCI-X-to-PCI-X bridge provides three slots, both with Extended Error Handling (EEH) for error recovering. All slots are PCI-X slots that operate at 133 MHz and 3.3 volt signaling. Figure 2-6 shows a conceptual diagram of the 7311-D20 I/O drawer.

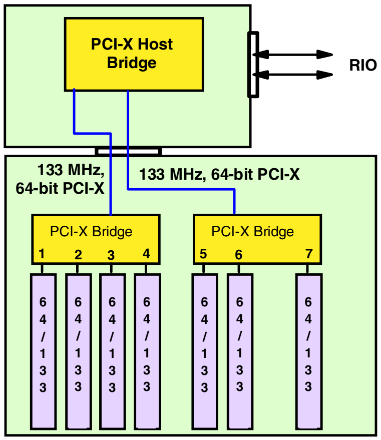


Figure 2-6 Conceptual diagram of the I/O drawer

2.6.3 I/O drawer cabling

The following sections provide additional information regarding cabling.

SPCN and RIO

For power control and monitoring of the I/O drawers, SPCN cables are used. The SPCN cables form a loop similar to the way the RIO cables do. The cabling starts from SPCN port 0 on the CEC to SPCN port 0 on the first I/O drawer connecting them from port 1 to port 0 of second I/O drawer or back to the CEC. For further information about SPCN refer to 3.4.6, “System Power Control Network (SPCN), power supplies, and cooling” on page 44.

The RIO cabling works in a similar way. The CEC connects from RIO port 0 to port 0 on the first I/O drawer. From port 1 on the I/O drawer a second I/O drawer could be connected on port 0. The last I/O drawer in the RIO loop connects from port 1 back to port 1 at the CEC. Figure 2-7 shows the cabling of the RIO cables as well as the cabling of SPCN.

The RIO cables used in Model 6C4 are different from the cables used in former systems; therefore, they could not be exchanged with cables from other systems, such as p660 or p680.

SPCN cables are the same as those used in former systems. The following RIO and SPCN cables are available:

- Remote I/O cable, 3.5 M (FC 3147)
- Remote I/O cable, 10 M (FC 3148)
- SPCN cable, 2 M (FC 6001)
- SPCN cable, 3 M (FC 6006)
- SPCN cable, 6 M (FC 6008)
- SPCN cable, 15 M (FC 6007)

Note: One 7311-D20 cannot be shared between two or more Models 6C4 at the same time. You can re-cable a drawer from one Model 6C4 to another Model 6C4 (see 2.6.1, “I/O drawer ID assignment” on page 23).

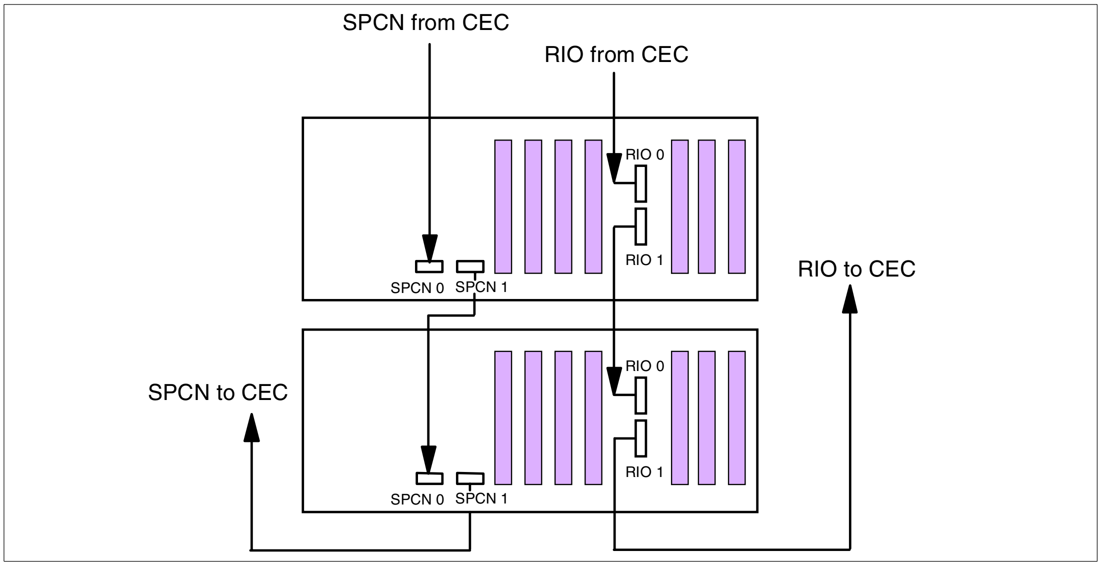


Figure 2-7 Remote I/O and SPCN cabling

SCSI internal cabling

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

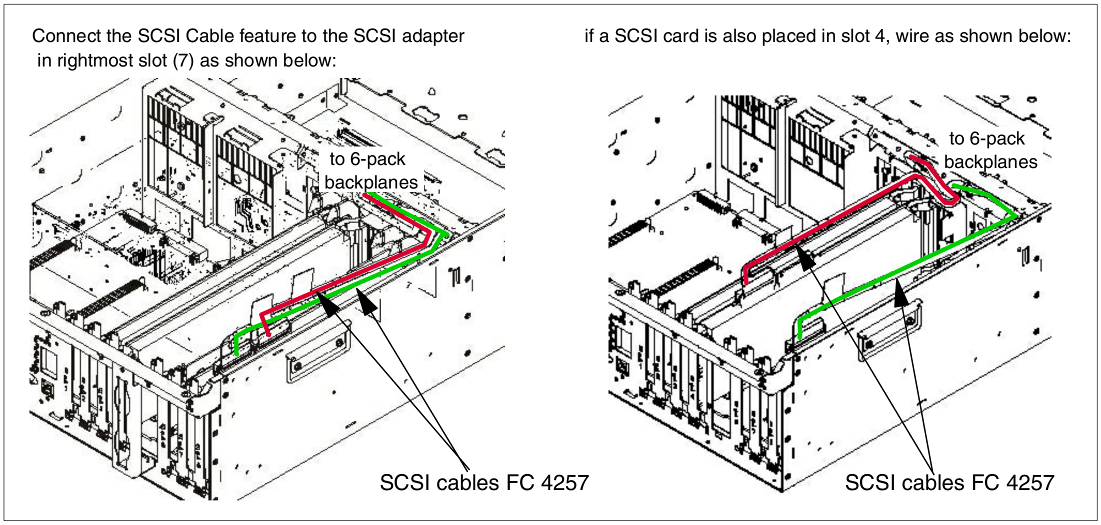
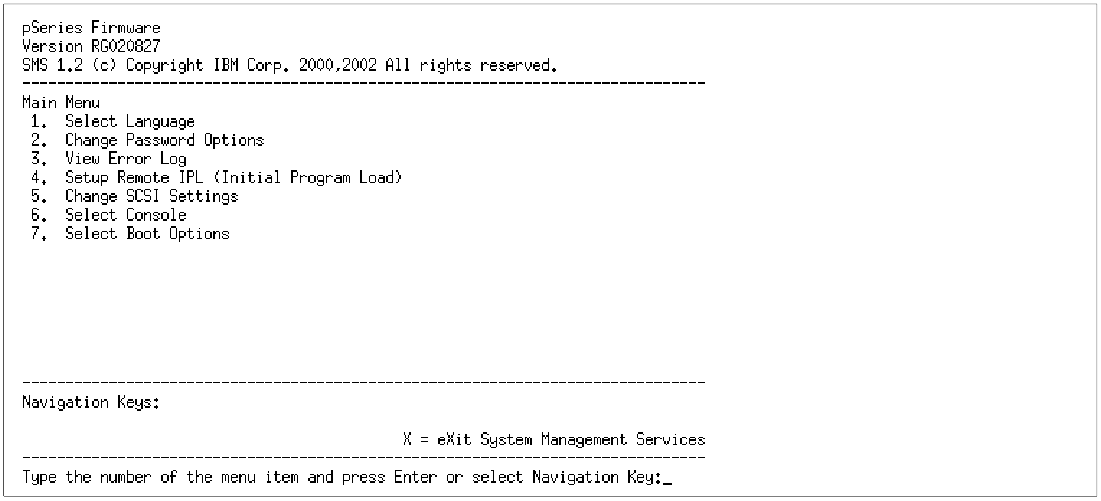

Connect the SCSI Cable feature to the SCSI adapter in rightmost slot (7) as shown below:
to 6-pack backplanes
SCSI cables FC 4257
if a SCSI card is also placed in slot 4, wire as shown below:
to 6-pack backplanes
SCSI cables FC 4257

Figure 2-8 Internal SCSI cabling

2.7 System Management Services (SMS)

When a p630 is equipped with either a graphic adapter connected to a graphics display, display, keyboard, and mouse device, or an ASCII display terminal connected to one of the first two system serial ports, you can use the System Management Services menus to view information about the system and perform tasks such as setting a password, changing the boot list, and setting the network parameters. Graphical SMS is not supported in LPAR mode.

To start the System Management Services, press the 1 key on the terminal, or in the graphic keyboard after the word `keyboard` appears and before the word `speaker` appears. After the text-based System Management Services start, the screen shown in Figure 2-9 displays.



```
pSeries Firmware
Version RG020827
SMS 1.2 (c) Copyright IBM Corp. 2000,2002 All rights reserved.
-------------------------------------------------------------------------------
Main Menu
 1.  Select Language
 2.  Change Password Options
 3.  View Error Log
 4.  Setup Remote IPL (Initial Program Load)
 5.  Change SCSI Settings
 6.  Select Console
 7.  Select Boot Options








-------------------------------------------------------------------------------
Navigation Keys:

                                             X = eXit System Management Services
-------------------------------------------------------------------------------
Type the number of the menu item and press Enter or select Navigation Key:_
```

Figure 2-9 System Management Services main menu

Note: The version of firmware currently installed in your system is displayed at the top of each screen. Processor and other device upgrades may require a specific version of firmware to be installed in your system.

On each menu screen, you are given the option of choosing a menu item and pressing Enter (if applicable), or selecting a navigation key. You can use the different options to set the password and protect our system set-up, review the error log for any kind of error reported during the first phases of the booting steps, or set-up the network environment parameters if you want the system boots from a NIM server. For more information see *IBM eServer pSeries 630 Model 6C4 and Model 6E4 Service Guide,* SA38-0604.

Use the Select Boot Options menu to view and set various options regarding the installation devices and boot devices:

1. Select Install or Boot a Device
2. Select Boot Devices
3. Multiboot Startup

Option 1 (Select Install or Boot a Device) allows you to select a device to boot from or install the operating system from. This selection is for the current boot only.

Option 2 (Select Boot Devices) allows you to set the boot list.

Option 3 (Multiboot Startup) toggles the multiboot startup flag, which controls whether the multiboot menu is invoked automatically on startup.

2.8 LPAR

LPAR stands for logical partitioning, and is the ability to divide a physical server into *virtual* logical servers each running a separate operating system. The Model 6C4 can be divided into up to four LPARs when combined with an I/O drawer. A 6-slot Model 6E4 supports up to three LPARs (a 4-slot Model 6E4 supports up to two LPARs). To better understand the requirements for LPAR, see 2.8.2, “LPAR minimum requirements” on page 29.

FC 9575 is required on new system orders to indicate the system is RIO and LPAR capable. For existing systems, FC 6575 adds RIO capability and FC 6576 adds LPAR capability to the 4-slot riser. FC 6575 and FC 6576 must be installed by a service representative.

Though it may not seem practical, running a machine with a single LPAR compared to full system partition (non-LPAR) mode provides for a faster system restart because the hypervisor has already provided some initialization, testing, and building of device trees. In environments where restart time is critical, it is recommended to test the single LPAR scenario to see if it meets the system recycle time objectives.

Depending on software installed on the p630, dynamic LPAR may be available or unavailable:

Dynamic LPAR available With dynamic LPAR available, the resources can be exchanged between partitions without stopping and rebooting the affected partitions. Dynamic LPAR requires AIX 5L Version 5.2 for all affected partitions and the HMC recovery software must be at Release 3 Version 1 (or higher). In partitions running AIX 5L Version 5.1 or Linux (see 2.10.2, “Linux” on page 33, for availability), the Dynamic Logical Partitioning menu is not available.

Dynamic LPAR unavailable Without dynamic LPAR, the resources in the partitions are static. Dynamic LPAR is unavailable for partitions running AIX 5L Version 5.1 or Linux (see Section 2.10.2, “Linux” on page 33 for availability). When you change or reconfigure your resource without dynamic LPAR, all the affected partitions be stopped and rebooted in order to make resource changes effective.

A server can contain a mix of partitions that support dynamic LPAR along with those that do not.

2.8.1 Hardware Management Console (HMC)

When the Models 6C4 and 6E4 are partitioned, an IBM Hardware Management Console for pSeries (HMC) is necessary. Either a dedicated 7315-C01 or an existing HMC from a p670 or p690 installation (FC 7316) can be used. If a Model 6C4 or 6E4 is only used in full system partition mode (no LPAR) outside a cluster, an HMC is not required. In this case, the Models 6C4 and 6E4 behave as non-partitionable pSeries models.

The HMC is a dedicated desktop workstation that provides a graphical user interface for configuring and operating pSeries servers functioning in either non-partitioned, LPAR, or clustered environments. It is configured with a set of hardware management applications for configuring and partitioning the server. One HMC is capable of controlling multiple pSeries servers. At the time this publication was written, a maximum of 16 non-clustered pSeries servers and a maximum of 64 LPARs are supported by one HMC.

The HMC provides two serial ports. One serial port should be used to attach a modem for the Service Agent (see 3.4.8, “Service Agent and Inventory Scout” on page 46, for details). The second port could be used to attach a server. If multiple servers should be attached to the HMC, additional serial ports are necessary. The ports could be provided by adding a maximum of two of the following features to the HMC:

- 8-Port Async Adapter (FC 2943)
- 128-Port Async Controller (FC 2944)

Note: To ensure that the Async adapter is installed into HMC and not in the server, make sure that the adapter is configured as a feature of the HMC at the time of order.

The HMC is connected with special attachment cables to the HMC ports of the Models 6C4 and 6E4. Only one serial connection to a server is necessary despite the number of LPARs. The following cables are available:

- FC 8121 Attachment Cable, HMC to host, 15 meters
- FC 8120 Attachment Cable, HMC to host, 6 meters

With these cables, the maximum length from any server to the HMC is limited to 15 meters. To extend this distance, a number of possibilities are available:

- Another HMC could be used for remote access. This remote HMC must have a network connection to the HMC, which is connected to the servers.
- AIX 5L Web-based System Manager Client could be used to connect to the HMC over the network or the Web-based System Manager PC client could be used, which runs on a Windows operating system-based or Linux operating system-based system.

- When a 128-Port Async Controller is used, the RS-422 cables connect to a RAN breakout box, which can be up to 330 meters. The breakout box is connected to the HMC port on the server using the attachment cable. When the 15-meter cable is used, the maximum distance the HMC can be is 345 meters, providing the entire cable length can be used.

The HMC provides a set of functions that are necessary to manage LPAR configurations. These functions include:

- Creating and storing LPAR profiles that define the processor, memory, and I/O resources allocated to an individual partition.
- Starting, stopping, and resetting a system partition.
- Booting a partition or system by selecting a profile.
- Displaying system and partition status.

  In a non-partitionable system, the LED codes are displayed in the operator panel. In a partitioned system, the operator panel shows the word `LPAR` instead of any partition LED codes. Therefore all LED codes for system partitions are displayed over the HMC.
- Virtual console for each partition or controlled system.

  With this feature, every LPAR can be accessed over the serial HMC connection to the server. This is a convenient feature when the LPAR is not reachable across the network or a remote NIM installation should be performed.

The HMC also provides a service focal point for the systems it controls. It is connected to the service processor of the system using the dedicated serial link, and must be connected to each LPAR using an Ethernet LAN for Service Focal Point and to coordinate dynamic LPAR operations. The HMC provides tools for problem determination and service support, such as call-home and error log notification through an analog phone line.

2.8.2 LPAR minimum requirements

Each LPAR must have a set of resources available. The minimum resources that are needed per LPAR (not per system) are the following:

- At least one processor per partition.
- At least 256 MB of physical memory per additional partition.
- At least one disk to store the operating system (for AIX, the rootvg).
- At least one disk adapter or integrated adapter to access the disk.
- At least one Ethernet adapter per partition to provide a network connection to the HMC, as well as general network access.
- A partition must have an installation method, such as NIM, and a means of running diagnostics, such as network diagnostics.

> **Note:** The minimum system memory required to run in LPAR mode with a single 256 MB partition is 1 GB.

2.8.3 Hardware guidelines for LPAR

There are a few limitations that should be considered when planning for LPAR, as discussed in the following sections.

Processor

There are no special considerations for processors. Each LPAR requires at least one processor.

Memory

Planning the memory for logical partitioning involves additional considerations to those already discussed in 2.2.1, “Memory options” on page 17. These considerations are different when using AIX 5L Version 5.1, AIX 5L Version 5.2, or Linux.

When a machine is in full system partition mode (no LPARs), all of the memory is dedicated to AIX 5L. When a machine is in LPAR mode, some of the memory used by AIX is relocated outside the AIX-defined memory range. In the case of a single small partition (256 MB), the first 256 MB of memory will be allocated to the hypervisor; 256 MB is allocated to translation control entries (TCEs) and to hypervisor per partition page tables; and 256 MB for the first page table for the first partition. TCE memory is used to translate the I/O addresses to system memory addresses. Additional small page tables for additional small partitions will fit in the page table block. Therefore, the memory allocated independently of AIX to create a single 256 MB partition is 1 GB.

With the previous memory statements in mind, LPAR requires at least 2 GB of memory for two or more LPARs on a p630. It is possible to create a single 256 MB LPAR partition on a 1 GB machine, however, this configuration should be used for validation of minimum configuration environments for test purposes only.

You must close any ISA or IDE device before any dynamic LPAR memory is removed from the partition that owns the ISA or IDE I/O. This includes the diskette drive, serial ports, CD-ROM, or DVD-ROM, for example.

The following rules apply for partitions with current service levels of AIX 5L or Linux:

- The minimum memory required to be assigned to a single LPAR is 256 MB. Additional memory can be configured in increments of 256 MB.
- The memory consumed outside AIX is from 0.75 GB up to 2 GB, depending on the amount of memory and the number of LPARs.
- With AIX 5L Version 5.1, partitions with a memory size larger than 16 GB, even if more than 16 GB of physical memory is installed, are not supported.
- With AIX 5L Version 5.2 and Linux, there are no limitations concerning partitions larger than 16 GB.

> **Note:** To create LPARs running AIX 5L Version 5.2 or Linux larger than 16 GB, the checkbox **Small Real Mode Address Region** must be checked (on the HMC, LPAR Profile, Memory Options dialog). Do not select this box if you are running AIX 5L Version 5.1.

I/O

The I/O devices are assigned on a slot level to the LPARs, meaning an adapter installed in a specific slot can only be assigned to one LPAR. If an adapter has multiple devices such as the 4-port Ethernet adapter or the Dual Ultra3 SCSI adapter, all devices are automatically assigned to one LPAR and cannot be shared.

The following I/O devices are connected over the same primary PCI-X bus to the PCI-X host bridge as the IDE devices (see Figure 2-1 for details). Therefore, all these ports must be assigned together to only one LPAR. These devices cannot be dynamically assigned:

- PCI slot 1 and slot 2
- All disks installed internally
- Internal SCSI port (located on the planar, not on the 6-slot riser card if equipped)
- External SCSI port
- Ethernet port U0.1-P1/E1
- IDE CD-ROM or IDE DVD-ROM
- Diskette drive
- Serial ports
- Keyboard and mouse

Note: The parallel port is not available in LPAR mode.

The I/O devices that you can independently assign are the second Ethernet port, slot 3, slot 4, slot 5, slot 6, the integrated SE SCSI adapter located on the 6-slot riser card (if equipped), and any I/O connected using the optional I/O drawer. These are also the devices that can be dynamically assigned, and may be assigned to different partitions.

The internal disks and the external SCSI port are driven by one SCSI chip on the I/O backplane. Therefore the internal disks and all external disks that are connected to the external SCSI port must be assigned together to the same LPAR and cannot be dynamically assigned.

If an internal IDE CD-ROM or IDE DVD-ROM is used to install several LPARs, complications may result because the IDE devices can only be assigned to one LPAR together with the internal SCSI disks and the disks attached to the external SCSI port. In this configuration, when a second LPAR is installed using the IDE device, these internal disks must be reassigned to whichever LPAR the IDE device uses. Therefore, you must be careful not to overwrite the disks of the first LPAR when using the internal IDE device to install another.

For servers with 4-slot riser cards, a possibility to provide access to CD-ROMs and DVD-RAMs for different LPARs is to use an externally attached DVD-RAM (FC 7210 Model 025) with a storage device enclosure (FC 7212 Model 102). This external DVD-RAM could be connected to a PCI SCSI adapter (FC 6203), which makes it easy to move the DVD-RAM between different LPARs. This solution provides an additional advantage of sharing this DVD-RAM between several servers by attaching it to SCSI adapters in different servers.

For servers with 6-slot riser cards, SCSI devices installed in the media bays, such as a DVD-RAM or tape device, can be assigned to any LPAR, and reassigned as required.

Every LPAR needs a disk for the operating system. Models 6C4 and 6E4 have up to four disks arranged in a 4-pack, which is connected to the internal SCSI port. As described previously, all SCSI devices (including all four disks) can only be assigned to one LPAR. Therefore, for additional LPARs external disk space is necessary.

For a Model 6C4, additional disk space could be provided by using up to two I/O drawers (7311-D20), which provide up to two 6-packs for each drawer. Each 6-pack requires a SCSI adapter (FC 6203 or FC 2498) for operation. This adapter provides two SCSI buses, which can attach to either internal or external devices. With these adapters, both 6-packs could be attached to only one adapter. For high availability reasons and for using each 6-pack in a different LPAR, two SCSI adapters are recommended.

For a Model 6E4, additional disk space could be provided by using a 2104 Expandable Storage Plus subsystem or a 7133 Serial Disk subsystem (SSA).

IBM 2104 Expandable Storage Plus

The IBM 2104 Expandable Storage Plus Model TU3 (tower) or DU3 (drawer) is a low-cost disk subsystem that supports up to 14 Ultra3 SCSI disks from 18.2 GB up to 146.8 GB at the time this publication was written. This subsystem could be used in splitbus mode, meaning the bus with 14 disks could be split into two buses with seven disks each. In this configuration two additional LPARs could be provided with up to seven disks for rootvg by using one Ultra3 SCSI adapter (FC 6203) for each LPAR.

For additional information on the IBM 2104 Expandable Storage Plus subsystem, visit the following Web site:

http://www.storage.ibm.com/hardsoft/products/expplus/expplus.htm

IBM 7133 Serial Disk Subsystem (SSA)

The IBM 7133 Serial Disk Subsystem Model T40 (tower) and Model D40 (rack-mount) provide a highly available storage subsystem for pSeries servers and is also a good solution for providing disks for booting additional LPARs. Disks are available from 18.2 GB up to 145.6 GB at the time this publication was written. SSA disk subsystems are connected to the server in loops. Each 7133 disk subsystem provides a maximum of four loops with a maximum of four disks each. Therefore up to four additional LPARs could be provided with disks for booting by using one Advanced Serial RAID Plus adapter (FC 6230) for each LPAR. Disk space for booting could be provided in Just a Bunch Of Disks (JBOD) or RAID mode.

> **Notes:** FC 6230 serial RAID adapters provide boot support from a RAID configured disk with firmware level 7000 and higher.
>
> Fastwrite cache must not be enabled on the boot resource SSA adapter.
>
> For more information on the SSA boot, see the SSA frequently asked questions located on the Web:
>
> http://www.storage.ibm.com/hardsoft/products/ssa/faq.html#microcode

For additional information about SSA, visit the following Web site:

http://www.storage.ibm.com/hardsoft/products/7133/7133.htm

2.9 Security

The Models 6C4 and 6E4 allow you to set two different types of passwords to limit the access to these systems. The *privileged access password* can be set from service processor menus or from System Management Services (SMS) utilities. It provides the user with the capability to access all service processor functions. This password is usually used by the system administrator or root user. The *general access password* can be set only from service processor menus. It provides limited access to service processor menus, and is usually available to all users who are allowed to power on the server, especially remotely.

2.10 Operating system requirements

The Models 6C4 and 6E4 are capable of running IBM AIX 5L for POWER and support appropriate versions of Linux. AIX 5L has been specifically developed and enhanced to exploit and support the extensive RAS features on IBM @server pSeries systems, and AIX 5L Version 5.2 supports dynamic logical partitioning.

2.10.1 AIX 5L

The Models 6C4 and 6E4 with LPAR support require AIX 5L Version 5.2 or AIX 5L Version 5.1 at Maintenance Level 3 (APAR IY32749). In order to boot from the CD, make sure you have one of the following media:

- AIX 5L Version 5.1 5765-E61, dated 10/2002 (CD# LCD4-1061-04) or later
- AIX 5L Version 5.2 5765-E62, initial CD-set (CD# LCD4-1133-00) or later

Note: For systems that are not used in LPAR mode, AIX 5L Version 5.1 at Maintenance Level 2, plus APAR IY31315, is the minimum requirement. In order to boot a non-LPAR system, the following CD is the minimum requirement: AIX 5L Version 5.1 5765-E61, dated 04/2002 (CD# LCD4-1061-03), although we recommend using the AIX versions mentioned previously.

AIX 5L Version 5.2 or later is required to support dynamic LPAR.

IBM periodically releases maintenance packages for the AIX 5L operating system. These packages are available on CD-ROM (FC 0907) or they can be downloaded from the Internet at:

`http://techsupport.services.ibm.com/server/fixes`

You can also get individual operating system fixes and information on how to obtain AIX 5L service at this site. If you have problems downloading the latest maintenance level, ask your IBM Business Partner or IBM representative for assistance.

To check your current AIX level enter the **`oslevel -r`** command. The output for AIX 5L Version 5.1 Maintenance Level 2 is `5100-02`.

AIX 5L application binary compatibility

IBM AIX 5L Version 5.2 preserves binary compatibility for 32-bit application binaries from previous levels of AIX Version 4 and AIX 5L, and for 64-bit applications compiled on previous levels of AIX 5L. 64-bit applications compiled on Version 4 must be recompiled to run on AIX 5L.

2.10.2 Linux

A Linux distribution, at the time this publication was written, is available through SuSE. For an overview of this support, see:

`http://www.ibm.com/servers/eserver/pseries/linux`

Full information on SuSE Linux Enterprise Server 8 for pSeries, see:

`http://www.suse.com/us/business/products/sles/index.html`

For all the latest in IBM Linux news, subscribe to the Linux Line. See:

`http://domino1.haw.ibm.com/linuxline`

Many of the features described in this document are operating system dependant and may not be available on Linux. For more information, please check:

http://www.ibm.com/servers/eserver/pseries/linux/whitepapers/linux_pseries.html

3

Availability, investment protection, expansion, and accessibility

The following sections provide more detailed information about configurations, upgrades, and design features that will help lower the total cost of ownership. This section assumes the benefits regarding AIX 5L. Support of these features using Linux may vary.

3.1 Autonomic computing

The IBM autonomic computing initiative is about using technology to manage technology. This initiative is an ongoing effort to create servers that respond to unexpected capacity demands and system glitches without human intervention. The goal: New highs in reliability, availability, and serviceability, and new lows in downtime and cost of ownership.

Today's pSeries offers some of the most advanced self-management features for UNIX servers on the market today.

Autonomic computing on IBM @server pSeries servers[7] describes the many self-configuring, self-healing, self-optimizing, and self-protecting features that are available on IBM @server pSeries servers.

Self-configuring

Autonomic computing provides self-configuration capabilities for the IT infrastructure. Today, IBM systems are designed to provide this at a feature level with capabilities like plug and play devices, and configuration setup wizards. Examples include:

- Virtual IP address (VIPA)
- IP multipath routing
- Microcode discovery services/inventory scout
- Hot-swappable disks
- Hot-plug PCI
- Wireless/pervasive system configuration
- TCP explicit congestion notification

Self-healing

For a system to be self-healing, it must be able to recover from a failing component by first detecting and isolating the failed component, taking it off-line, fixing or isolating the failed component, and reintroducing the fixed or replacement component into service without any application disruption. Examples include:

- Multiple default gateways
- Automatic system hang recovery
- Automatic dump analysis and e-mail forwarding
- EtherChannel automatic failover
- Graceful processor failure detection and failover
- HACMP and HAGeo
- First failure data capture
- Chipkill™ ECC Memory, dynamic bit-steering
- Memory scrubbing
- Automatic, dynamic deallocation (processors, LPAR, PCI buses/slots)
- Electronic Service Agent - *call-home* support

[7] http://www-3.ibm.com/autonomic/index.shtml

Self-optimization

Self-optimization requires a system to efficiently maximize resource utilization to meet the end-user needs with no human intervention required. Examples include:

- Static LPAR
- Dynamic LPAR
- Workload manager enhancement
- Extended memory allocator
- Reliable, scalable cluster technology (RSCT)
- PSSP cluster management and Cluster Systems Management (CSM)

Self-protecting

Self-protecting systems provide the ability to define and manage the access from users to all the resources within the enterprise, protect against unauthorized resource access, detect intrusions and report these activities as they occur, and provide backup/recovery capabilities that are as secure as the original resource management systems. Examples include:

- Kerberos V5 Authentication (authenticates requests for service in a network)
- Self-protecting kernel
- LDAP directory integration (LDAP aids in the location of network resources)
- SSL (manages Internet transmission security)
- DigitalCertificates
- Encryption (prevents unauthorized use of data)

3.2 High-availability solution

For even greater availability and reliability, the Models 6C4 and 6E4 support IBM High Availability Cluster Multiprocessing (HACMP) software clustering solution. This solution, when combined with applications that meet IBM ClusterProven® standards, provides an excellent base for high availability, an essential ingredient of e-commerce.

The Models 6C4 and 6E4 logically have three serial ports, that is, front S1, rear S1, S2, and S3. The service processor menu is only shown on either S1 or S2 port if an ASCII terminal is connected to the port. It is recommended that HACMP or UPS functions use the S3 port. However, the need may arise to use HACMP and UPS at the same time. For only such a demand, the S2 port also can be used for HACMP. When the machine is in the standby state, the service processor is looking at both the S1 and S2 ports to see if any character is coming in from either port. If the user types any key from the ASCII terminal and that character comes in, the service processor selects that port to show the service processor menu to the ASCII terminal. In order to prevent HACMP traffic from appearing as a user key press, the service processor watches if the first character coming in from the port is Ctrl+D or not, because HACMP initially sends out Ctrl+D code to declare that the port is used by HACMP, and if Ctrl+D code is coming in from the S2 port, then the service processor disables the service processor menu for the S2 port. This menu disablement lasts until the next time the service processor is reset, either by the pin-hole reset switch or by service processor setup menu operation. On the other hand, there is no such indicator for the UPS. Therefore, if you want to use HACMP and UPS at the same time, use the S3 port for UPS and S2 port for HACMP. Also, once you have set up this configuration, decided to stop using a UPS, HACMP, and would like to use the service processor menu on S2 port, the service processor reset must be initiated.

Note: Serial ports 2 and 3 support HACMP heartbeat functions. If the S2 is used for HACMP heartbeat, then it cannot be used again for the ASCII terminal attached to the service processor until the service processor is reset.

Order FC 3124 (HACMP serial-to-serial cable - drawer-to-drawer 3.7 meter) or FC 3125 (HACMP serial to serial cable - rack to rack 8 meter) for the serial non-IP heartbeat connections. FC 3925 converters are required for each end of either cable to attach it to the system.

3.3 IBM @server Cluster 1600 and SP switch attachment

The Model 6C4 is supported in either a non-switched IBM @server Cluster 1600 or a switched Cluster 1600 system using the SP Switch2 adapter (FC 8397). p630 servers may also function as a control workstation (CWS).

A Cluster 1600 can scale up to 128 servers or 128 operating systems using PSSP 3.5 on AIX 5L Version 5.1 or CSM[8] 1.3 on AIX 5L Version 5.2. The cluster management server can be running on any server or LPAR running AIX 5L Version 5.2 and CSM 1.3. CSM supports a non-switched environment only. PSSP[9] Version 3.4 or Version 3.5 is required to support SP Switch2 adapter (FC 8397) with AIX 5L Version 5.1. The SP Switch2 adapter (FC 8397) must be placed in adapter slot 3.

IBM intends to provide support for AIX 5L Version 5.2 with the PSSP for AIX product on Cluster 1600 systems in 2003. See 1.5, “Statements of direction” on page 12 for more information.

To attach a Model 6C4 to a Cluster 1600, an HMC is required. Either a dedicated 7315-C01 can be used or an existing HMC from a p670 or p690 installation. One HMC can also control several Model 6C4s that are part of the cluster. If a Model 6C4 configured in LPAR mode is part of the cluster, all LPARs must be part of the cluster. It is not possible to use selected LPARs as part of the cluster and use others for non-cluster use.

The HMC uses a dedicated connection to the Model 6C4 to provide the functions needed to control the server, such as powering the system on and off. The HMC must have an Ethernet connection to the CWS. Each LPAR in Model 6C4 must have an Ethernet adapter to connect to the CWS *trusted* LAN.

Information regarding HMC control of clustered servers under the control of IBM Parallel Systems Support Programs for AIX (PSSP) or Cluster Systems Management for AIX (CSM) can be found in the Scaling Statement section of the Family 9078+01 IBM @server Cluster 1600, 9078-160 sales manual, accessible on IBMlink:

http://www.ibmlink.ibm.com

3.4 Reliability, availability, and serviceability (RAS) features

Excellent quality and reliability are inherent in all aspects of the Models 6C4 and 6E4 design and manufacture, and the fundamental principle of the design approach is to minimize outages. The RAS features help to ensure that the systems operates when required, performs reliably, and efficiently handles any failures that may occur. This is achieved using capabilities provided by both the hardware and the AIX 5L operating system.

[8] Cluster Systems Management (CSM)

[9] Parallel System Support Programs (PSSP)

Mainframe-class diagnostic capability based on internal error checkers, First-Failure Data Capture (FFDC), and run-time analysis is provided. This monitoring of all internal error check states is provided for processor, memory, I/O, power, and cooling components, and is aimed at eliminating the need to try to recreate failures later for diagnostic purposes. The unique IBM RAS capabilities are important for the availability of your server.

The following features provide the Models 6C4 and 6E4 with UNIX industry-leading RAS:

- Fault avoidance through highly reliable component selection, component minimization, and error handling technology designed into the chips.
- Improved reliability through processor operation at a lower voltage enabled by the use of copper chip circuitry and Silicon-on-Insulator technology.
- Fault tolerance with redundancy, dual line cords, and concurrent maintenance for power and cooling (using optional redundant hot-swap power supplies and fans).
- Automatic First Failure Data Capture (FFDC) and diagnostic fault isolation capabilities.
- Concurrent run-time diagnostics based on First Failure Data Capture.
- Predictive failure analysis on processors, caches, memory, and disk drives.
- Dynamic error recovery.
- Error Checking and Correction (ECC) or equivalent protection (such as refetch) on main storage, all cache levels (1, 2, and 3), and internal processor arrays.
- Dynamic processor deallocation based on run-time errors.
- Persistent processor deallocation (boot-time deallocation based on run-time errors).
- Persistent deallocation extended to memory.
- Chipkill correction in memory.
- Memory scrubbing and redundant bit-steering for self-healing.
- Industry-leading PCI bus parity error recovery as first introduced on the p690 systems.
- Hot-plug functionality of the PCI bus I/O subsystem.
- PCI bus and slot deallocation.
- Disk drive fault tracking.
- Avoiding checkstops with process error containment.
- Environmental monitoring (temperature and power supply).
- Auto-reboot.
- Disk mirroring (RAID1) and disk controller duplexing capability are provided by the AIX operating system.

Some of the RAS features of the p630 are covered in more detail in the following sections.

3.4.1 Service processor

The converged service processor is a specialized device that is situated on the system board and provides a number of different functions, discussed below:

- With the machine powered off, the service processor is in an idle state waiting for either a power-on command or a keystroke from any of the TTYs attached to either of the S1 or S2 ports. At this point, `OK` is displayed on the control panel LED.
- Immediately after power on, the SPCN, a function of the service processor, controls the powering up of all devices needed during the boot process. When the SPCN has completed its tasks, the CSP, using an onboard processor, checks for CPU and memory

resources and then tests them. After the CPU and memory tests have completed, the service processor then hands the rest of the boot process over to system firmware. This changeover occurs when the 9*xxx* LED codes become E*xxx* codes.

- With AIX in control of the machine, the service processor is still working and checking the system for errors. Also, the surveillance function of the service processor is monitoring AIX to check that it is still running and has not stalled.
- With the machine powered off but power still attached (standby), any TTY keyboard attached to either the S1 or S2 native serial port having the Enter key depressed will cause the service processor to display the service processor main menu. This menu and subsequent menus will be discussed in the following sections.

Service processor main menu

The service processor main menu and subsequent menus are only visible on an ASCII screen attached to the native serial ports. Example 3-1 shows the service processor main menu.

Example 3-1 Service processor main menu

```
Service Processor Firmware
       Version: RR020516
 Copyright 2001, IBM Corporation
            100FB5A
________________________________
            MAIN MENU

 1. Service Processor Setup Menu
 2. System Power Control Menu
 3. System Information Menu
 4. Language Selection Menu
 5. Call-In/Call-Out Setup Menu
 6. Set System Name
99. Exit from Menus
```

Useful information contained within this menu is defined as follows:

RR020516 Service processor firmware level and date code. The RR020516 firmware indicated in this example is a preproduction level. Generally available versions will reflect a later date code.

100FB5A Machine serial number.

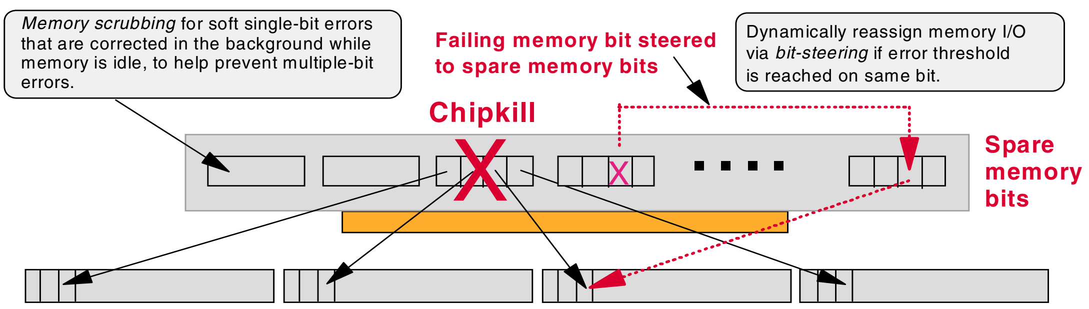
Service Processor Setup menu These menus allow you to set passwords to provide added security to your system and update machine firmware.

System Power Control menu These menus allow you to control some aspects of how the machine powers on and how much of the boot process you wish to complete.

System Information menu From this menu option, information about the previous boot sequence and errors produced can be viewed. Additionally, this menu provides a means of checking and altering the availability of CPUs and memory.

3.4.2 Memory reliability, fault tolerance, and integrity

The Models 6C4 and 6E4 use Error Checking and Correcting (ECC) circuitry for system memory to correct single-bit and to detect double-bit memory failures. Detection of double-bit memory failures helps maintain data integrity. Furthermore, the memory chips are organized such that the failure of any specific memory module only affects a single bit within a four bit ECC word (*bit-scattering*), thus allowing for error correction and continued operation in the presence of a complete chip failure (*Chipkill recovery*). The memory DIMMs also utilize *memory scrubbing* and thresholding to determine when spare memory modules within each bank of memory should be used to replace ones that have exceeded their threshold of error count (*dynamic bit-steering*). Memory scrubbing is the process of reading the contents of the memory during idle time and checking and correcting any single-bit errors that have accumulated by passing the data through the ECC logic. This function is a hardware function on the memory controller chip and does not influence normal system memory performance.



Memory scrubbing for soft single-bit errors that are corrected in the background while memory is idle, to help prevent multiple-bit errors.

Dynamically reassign memory I/O via *bit-steering* if error threshold is reached on same bit.

Scatter memory chip bits across four separate EC words for Chipkill recovery

- ***Bit-scattering*** allows normal single bit ECC error processing, thereby keeping system running with a ***Chipkill*** failure.
- ***Bit-steering*** allows memory lines from a spare memory chip to be dynamically reassigned to a memory module with a faulty line to keep the system running.

Figure 3-1 Main storage ECC and extensions

3.4.3 First failure data capture, diagnostics, and recovery

If a problem should occur, the ability to correctly diagnose it is a fundamental requirement upon which improved availability is based. The Models 6C4 and 6E4 incorporate un-matched capability in start-up diagnostics (see 3.4.1, "Service processor" on page 39) and in run-time First Failure Data Capture based on strategic error checkers built into the chips.

Any errors detected by the pervasive error checkers are captured into Fault Isolation Registers (FIRs), which can be interrogated by the service processor. The service processor in the Models 6C4 and 6E4 has the capability to access system components using special purpose service processor ports or by access to the error registers (see Figure 3-2 on page 42).

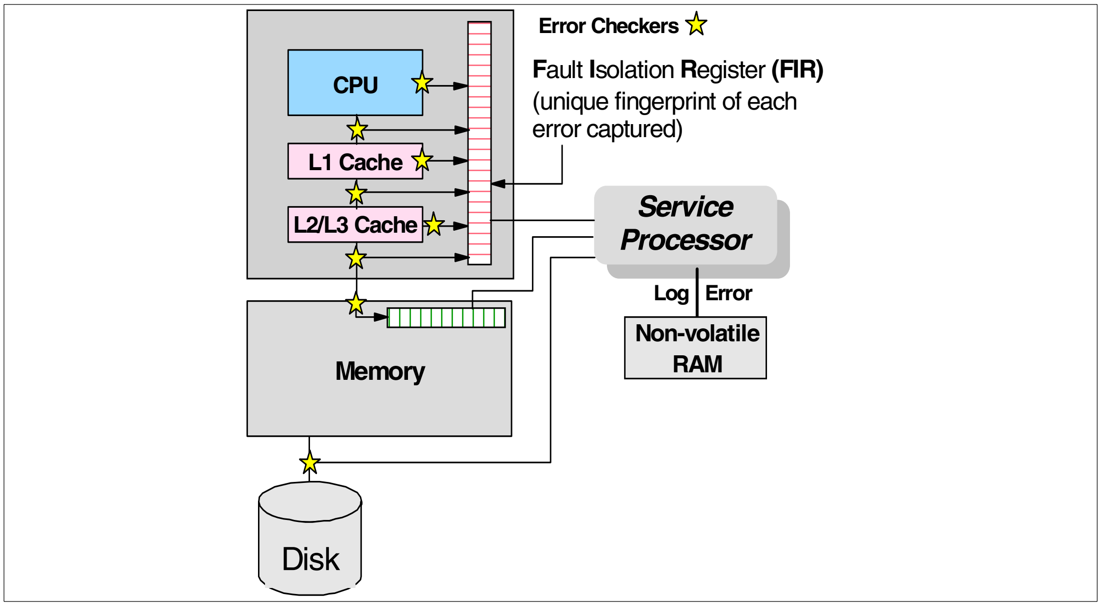


Figure 3-2 Fault Isolation Register

The FIRs are important because they enable an error to be uniquely identified, thus enabling the appropriate action to be taken. Appropriate actions might include such things as a bus retry, ECC correction, or system firmware recovery routines. Recovery routines could include dynamic deallocation of potentially failing components such as a processor or L2 cache.

Errors are logged into the system non-volatile random access memory (NVRAM) and the service processor event history log, along with a notification of the event to AIX for capture in the operating system error log. Diagnostic Error Log Analysis (diagela) routines analyze the error log entries and invoke a suitable action such as issuing a warning message. If the error can be recovered, or after suitable maintenance, the service processor resets the FIRs so that they can accurately record any future errors.

The ability to correctly diagnose any pending or firm errors is a key requirement before any dynamic or persistent component deallocation or any other reconfiguration can take place.

3.4.4 Dynamic or persistent deallocation

Dynamic deallocation of potentially failing components is non-disruptive, allowing the system to continue to run. Persistent deallocation occurs when a failed component is detected and is then deactivated at the subsequent boot time.

Dynamic deallocation functions include:

- Processor
- L3 cache line delete
- PCI bus and slots

For dynamic processor deallocation, the service processor performs a predictive failure analysis based on any recoverable processor errors that have been recorded. If these transient errors exceed a defined threshold, the event is logged and the processor is deallocated from the system while the operating system continues to run. This feature (named *CPU guard*) enables maintenance to be deferred until a suitable time. Processor deallocation can only occur if there are sufficient functional processors.

To verify whether cpuguard has been enabled, run the following command:

```
lsattr -El sys0 | grep cpuguard
```

If enabled, the output will be similar to the following:

```
cpuguard    enable      CPU Guard     True
```

If the output shows cpuguard as disabled, enter the following command to enable it:

```
chdev -l sys0 -a cpuguard='enable'
```

> **Note:** The use of cpuguard is only effective on systems with three or more functional processors on AIX 5L Version 5.1, or two or more with AIX 5L Version 5.2.

Cache or cache-line deallocation is aimed at performing dynamic reconfiguration to bypass potentially failing components. This capability is provided for both L2 and L3 caches. The L1 data cache and L2 data and directory caches can provide dynamic detection and correction of hard or soft array cell failures. Dynamic run-time deconfiguration is provided if a threshold of L1 or L2 recovered errors is exceeded.

In the case of an L3 cache run-time array single-bit solid error, the spare chip resources are used to perform a line delete on the failing line.

System bus recovery (retry) is provided for any address or data parity errors on the GX bus or for any address parity errors on the fabric bus.

PCI hot-plug slot fault tracking helps prevent slot errors from causing a system machine check interrupt and subsequent reboot. This provides superior fault isolation and the error affects only the single adapter. Run time errors on the PCI bus caused by failing adapters will result in recovery action. If this is unsuccessful, the PCI device will be gracefully shut down. Parity errors on the PCI bus itself will result in bus retry and, if uncorrected, the bus and any I/O adapters or devices on that bus will be deconfigured.

The Models 6C4 and 6E4 support PCI Extended Error Handling (EEH) if it is supported by the PCI adapter. In the past, PCI bus parity errors caused a global machine check interrupt, which eventually required a system reboot in order to continue. In the Model 6C4 and 6E4 systems, new hardware, system firmware, and AIX interaction has been designed to allow transparent recovery of intermittent PCI bus parity errors, and graceful transition to the I/O device available state in the case of a permanent parity error in the PCI bus.

EEH-enabled adapters respond to a special data packet generated from the affected PCI slot hardware by calling system firmware, which will examine the affected bus, allow the device driver to reset it, and continue without a system reboot.

Persistent deallocation functions include:

- Processor
- Memory
- Deconfigure or bypass failing I/O adapters

Following a hardware error that has been flagged by the service processor, the subsequent reboot of the system will invoke extended diagnostics. If a processor or L3 cache has been marked for deconfiguration by persistent processor deallocation, the boot process will attempt to proceed to completion with the faulty device automatically deconfigured. Failing I/O adapters will be deconfigured or bypassed during the boot process.

The auto-restart (reboot) option, when enabled, can reboot the system automatically following an unrecoverable software error, software hang, hardware failure, or environmentally-induced failure (such as loss of power supply).

3.4.5 UE-Gard

The UE-Gard (Uncorrectable Error-Gard) is a RAS feature that enables AIX 5L Version 5.2 in conjunction with hardware and firmware support to isolate certain errors that would previously have resulted in a condition where the system had to be stopped (checkstop condition). The isolated error is being analyzed to determine if AIX can terminate the process that suffers the hardware data error instead of terminating the entire system.

UE-Gard is not to be confused with (dynamic) CPU Guard. CPU Guard takes a CPU dynamically offline after a threshold of recoverable errors is exceeded, to avoid system outages.

For memory errors the firmware will analyze the severity and record it in a RTAS[10] log. AIX will be called from firmware with a pointer to the log. AIX will analyze the log to determine whether the error is recoverable. If the error is recoverable then AIX will resume. If the error is not fully recoverable then AIX will determine whether the process with the error is critical. If the process is not critical then it will be terminated by issuing a SIGBUS signal with an UE siginfo indicator. In the case where the process is a critical process then the system will be terminated as a machine check problem.

3.4.6 System Power Control Network (SPCN), power supplies, and cooling

Environmental monitoring related to power, fan operation, and temperature is done by the System Power Control Network. Critical power events, such as a power loss, trigger appropriate signals from hardware to impacted components to assist in the prevention of data loss without operating system intervention. Non-critical environmental events are logged and reported to the operating system.

A SYSTEM_HALT warning is issued to the operating system if the inlet air temperature rises above a preset maximum limit, or if two or more system fan units are slow or stopped. This warning will result in an immediate system shutdown action.

Hot-plug power supplies

The Models 6C4 and 6E4 can be configured with an additional power supply to provide redundancy should a failing unit need to be replaced. When configured with redundant power supplies, the failed power supply can be replaced concurrently and with minimum disruption. Note that when ambient temperatures exceed 32 C (92 F), it is advisable to shut down the machine prior to replacing the faulty power supply. When the additional power supply is ordered, a redundant processor cooling fan (FC 6557) is required. Each power supply contains two integrated cooling fans.

Hot-plug fans

Both Models 6C4 and 6E4 have a number of fans that can be changed concurrently. To assist in the identification of a failing fan, each unit is equipped with an amber LED that will illuminate when a fault is detected.

Processor cooling fan

The processors are cooled with either one or two variable-speed cooling fans arranged in tandem, with the second fan being optional (FC 6557). If the machine is configured with two

[10] RunTime Abstraction Services (RTAS) is the local firmware that is replicated to each LPAR of the system.

processor cooling fans, then either fan can provide redundancy in the event of a single fan failure (the remaining processor cooling fan will increase speed when required). Any failed fan can be replaced concurrently therefore eliminating the need for system downtime.

PCI adapter cooling fans

The PCI adapters are cooled by two banks of two fans placed side-by-side. These fans draw air in from the front of the machine, passing it across the SCSI disks, then blowing it out across the PCI adapters. Each of these pairs of fans can be changed concurrently. Any concurrent change must be accomplished within a five-minute time span or the machine will power down.

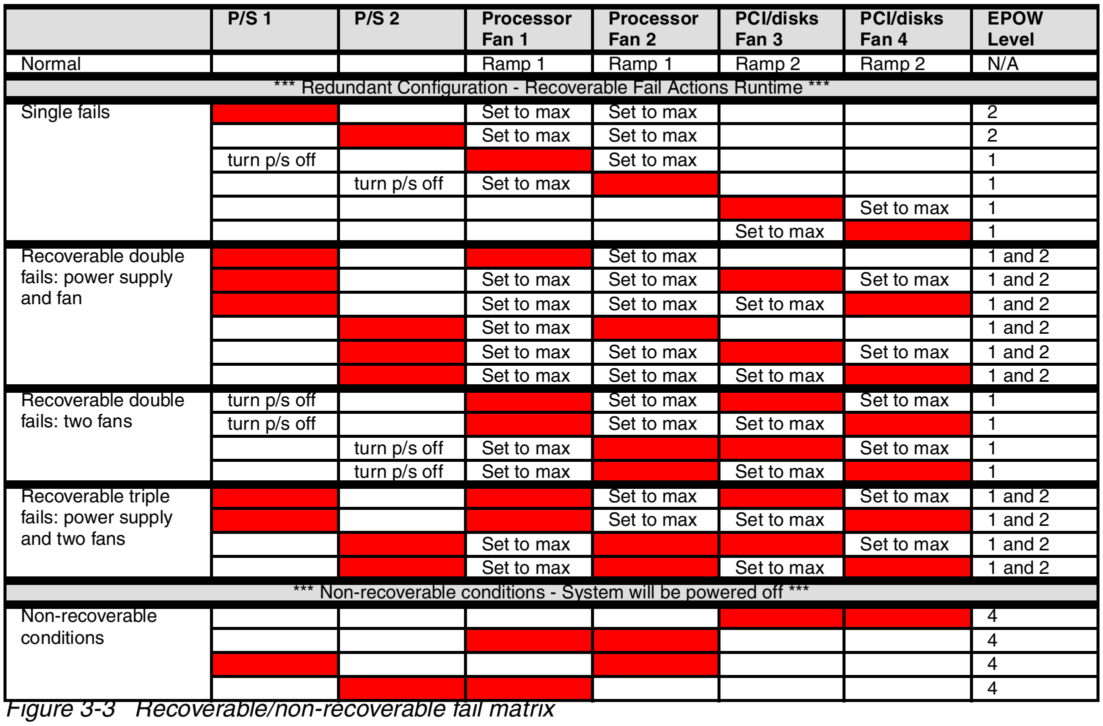
3.4.7 Early Power-Off Warning (EPOW)

Both critical and non-critical power supply and fan failures generate a signal that SPCN reports to AIX through the service processor interface as an Early Power On Warning error message.

The following is a summary of the p630 EPOW functions:

- EPOW 1 - Warn Cooling: This type of fault occurs when one of the system fans is not working. SPCN sends an alert to CSP and CSP flags for Event_Scan to pick up and place an entry in the error log.
- EPOW 2 - Warn Power: This type of fault occurs when one of the system's redundant supplies stops working. SPCN sends an alert to CSP and CSP flags for Event_Scan to pick up and place entry in the error log.
- EPOW 4 - System Halt (shutdown within 20 seconds): This type of EPOW is flagged if the system must shut down for thermal reasons. This happens when SPCN detects that two or more fans are not performing (or missing). For this type of message, SPCN powers off the offending domain if AIX does not power-off the system first.

There are two independent cooling domains in the Models 6C4 and 6E4, a top and a bottom section. As such, some multiple fan fail conditions are not critical. Figure 3-3 on page 46 describes how the power and cooling redundancy works for the Models 6C4 and 6E4. You can match the EPOW level generated by the failures with the events reported inside the table.



	P/S 1	P/S 2	Processor Fan 1	Processor Fan 2	PCI/disks Fan 3	PCI/disks Fan 4	EPOW Level
Normal			Ramp 1	Ramp 1	Ramp 2	Ramp 2	N/A
*** Redundant Configuration - Recoverable Fail Actions Runtime ***							
Single fails	🟥		Set to max	Set to max			2
		🟥	Set to max	Set to max			2
	turn p/s off		🟥	Set to max			1
		turn p/s off	Set to max	🟥			1
					🟥	Set to max	1
					Set to max	🟥	1
Recoverable double fails: power supply and fan	🟥		🟥	Set to max			1 and 2
	🟥		Set to max	Set to max	🟥	Set to max	1 and 2
	🟥		Set to max	Set to max	Set to max	🟥	1 and 2
		🟥	Set to max	🟥			1 and 2
		🟥	Set to max	Set to max	🟥	Set to max	1 and 2
		🟥	Set to max	Set to max	Set to max	🟥	1 and 2
Recoverable double fails: two fans	turn p/s off		🟥	Set to max	🟥	Set to max	1
	turn p/s off		🟥	Set to max	Set to max	🟥	1
		turn p/s off	Set to max	🟥	🟥	Set to max	1
		turn p/s off	Set to max	🟥	Set to max	🟥	1
Recoverable triple fails: power supply and two fans	🟥		🟥	Set to max	🟥	Set to max	1 and 2
	🟥		🟥	Set to max	Set to max	🟥	1 and 2
		🟥	Set to max	🟥	🟥	Set to max	1 and 2
		🟥	Set to max	🟥	Set to max	🟥	1 and 2
*** Non-recoverable conditions - System will be powered off ***							
Non-recoverable conditions					🟥	🟥	4
			🟥	🟥			4
	🟥			🟥			4
		🟥	🟥				4

Figure 3-3 Recoverable/non-recoverable fail matrix

A white box indicates a device that is present and working correctly. A painted box indicates a failed or not present resource.

> **Note:** When the EPOW level is 4, hardware responds as the description (shut down within few seconds). However, the error is logged and reported to the AIX error log as an EPOW 2.

3.4.8 Service Agent and Inventory Scout

Service Agent and Inventory Scout are two tools that can be used on the Models 6C4 and 6E4 to enable you to maintain the maximum availability of your system. Each item performs a different task, as discussed below:

Service Agent Service Agent is the successor to Service Director. It is a process that monitors the machine for problems requiring a service activity. If any problems are detected, then Service Agent will alert the IBM service organization to request service and, if required, will send a notification by e-mail to the system administrator, for example. Along with the request for service, sense data is also transmitted to enable an action plan to be prepared by the product support specialists. All data to be sent can be monitored by you before sending, if required.

This tool will be set up by your Customer Engineer upon your request.

Inventory Scout Inventory Scout is a tool that can be downloaded from the Internet. This tool will enable you to check the firmware or microcode levels of all of the devices in your system and advise you as to which code levels require attention. Inventory Scout is also being shipped with AIX 5L as a fileset. For more information, view the following Web page:

`http://techsupport.services.ibm.com/server/aix.invscoutMDS`

The option for the electronic service agent, or the service processor, to place a call to IBM is available at no additional cost, provided the system is under warranty or an IBM service or maintenance contract is in place. The electronic service agent monitors and analyzes system errors. For non-critical errors, service agent can place a service call automatically to IBM without customer intervention. For critical system failures, the dial-out is performed by the service processor itself, which also has the ability to send out an alert automatically using the telephone line to dial a paging service. This function is set up and controlled by the customer, not by IBM. It is not enabled by default. A hardware fault will also turn on the two attention indicators (one is located on the front and the other is located on the rear of the system) to alert the user of a hardware problem.

Related publications

The publications listed in this section are considered particularly suitable for a more detailed discussion of the topics covered in this Redpaper.

IBM Redbooks

For information on ordering these publications, see “How to get IBM Redbooks” on page 51.

- *AIX Logical Volume Manager from A to Z: Introduction and Concepts,* SG24-5432
- *AIX Logical Volume Manager from A to Z: Troubleshooting and Commands,* SG24-5433
- *IBM @server pSeries 690 System Handbook,* SG24-7040
- *Practical Guide for SAN with pSeries,* SG24-6050
- *Problem Solving and Troubleshooting in AIX 5L*, SG24-5496
- *Understanding IBM @server pSeries Performance and Sizing,* SG24-4810

Other resources

These publications are also relevant as further information sources:

- *7014 Series Model T00 and T42 Rack Installation and Service Guide*, SA38-0577, contains information regarding the 7014 Model T00 and T42 Rack, in which this server may be installed.
- *Flat Panel Display Installation and Service Guide*, SA23-1243, contains information regarding the 7316-TF2 Flat Panel Display, which may be installed in your rack to manage your system units.
- *IBM @server pSeries 630 Model 6C4 and Model 6E4 Installation Guide*, SA38-0605, contains detailed information on installation, cabling, and verifying server operation.
- *IBM @server pSeries 630 Model 6C4 and Model 6E4 Service Guide*, SA38-0604, contains reference information, maintenance analysis procedures (MAPs), error codes, removal and replacement procedures, and a parts catalog.
- *IBM @server pSeries 630 Model 6C4 and Model 6E4 User’s Guide*, SA38-0606, contains information to help users use the system, use the service aids, and solve minor problems.
- *RS/6000 Adapters, Devices, and Cable Information for Multiple Bus Systems*, SA38-0516, contains information about adapters, devices, and cables for your system. This manual is intended to supplement the service information found in the Diagnostic Information for Multiple Bus Systems documentation.
- *RS/6000 and eServer pSeries Diagnostics Information for Multiple Bus Systems*, SA38-0509, contains diagnostic information, service request numbers (SRNs), and failing function codes (FFCs).
- *RS/6000 and pSeries PCI Adapter Placement Reference*, SA38-0538, contains information regarding slot restrictions for adapters that can be used in this system.
- *System Unit Safety Information*, SA23-2652, contains translations of safety information used throughout the system documentation.

Referenced Web sites

These Web sites are also relevant as further information sources:

- AIX 5L operating system maintenance packages downloads

  `http://techsupport.services.ibm.com/server/fixes`
- Autonomic computing on IBM eServer pSeries servers

  `http://www-3.ibm.com/autonomic/index.shtml`
- Ceramic Column Grid Array (CCGA), see IBM Chip Packaging

  `http://www.ibm.com/chips/micronews`
- Copper circuitry

  `http://www-3.ibm.com/chips/bluelogic/showcase/copper/`
- Frequently asked SSA-related questions

  `http://www.storage.ibm.com/hardsoft/products/ssa/faq.html`
- Hardware documentation

  `http://www.ibm.com/servers/eserver/pseries/library/hardware_docs`
- IBM @server support: Fixes

  `http://techsupport.services.ibm.com/server/fixes`
- IBM @server support: Tips for AIX administrators

  `http://techsupport.services.ibm.com/server/aix.techTips`
- IBM Linux news - subscribe to the Linux Line

  `https://www6.software.ibm.com/reg/linux/linuxline-i`
- IBM online sales manual

  `http://www.ibmlink.ibm.com`
- Linux for IBM @server pSeries

  `http://www-1.ibm.com/servers/eserver/pseries/linux/`
- Microcode discovery service

  `http://techsupport.services.ibm.com/server/aix.invscoutMDS`
- Pervasive system management

  `http://www.ibm.com/servers/pervasivesm/`
- POWER4 system microarchitecture - comprehensively described in the IBM Journal of Research and Development, Vol 46 No.1 January 2002

  `http://www.research.ibm.com/journal/rd46-1.html`
- PowerPC Microprocessor Common Hardware Reference Platform (CHRP): A System Architecture

  `http://www.mkp.com/books_catalog/catalog.asp?ISBN=1-55860-394-8`
- SCSI T10 Technical Committee,

  `http://www.t10.org`
- Silicon on Insulator (SOI) technology

  `http://www-3.ibm.com/chips/bluelogic/showcase/soi`
- SSA boot FAQ

  `http://www.storage.ibm.com/hardsoft/products/ssa/faq.html#microcode`

- SuSE Linux Enterprise Server 8 for pSeries information

  http://www.suse.de/us/business/products/server/sles/i_pseries.html

How to get IBM Redbooks

You can order hardcopy Redbooks, as well as view, download, or search for Redbooks at the following Web site:

ibm.com/redbooks

You can also download additional materials (code samples or diskette/CD-ROM images) from that site.

IBM Redbooks collections

Redbooks are also available on CD-ROMs. Click the CD-ROMs button on the Redbooks Web site for information about all the CD-ROMs offered, as well as updates and formats.

IBM®

pSeries 630 Models 6C4 and 6E4 Technical Overview and Introduction

Redpaper

Logical partitionable, I/O drawer expandable

Two unique models: Deskside/desktop or rack-mount

High-end reliability, availability, and serviceability features

This document provides a comprehensive guide covering IBM @server pSeries 630 Models 6C4 and 6E4 servers. Major hardware offerings are introduced and their prominent functions discussed.

Professionals wishing to acquire a better understanding of IBM @server pSeries products may consider reading this document. The intended audience includes:

- Customers
- Sales and marketing professionals
- Technical support professionals
- IBM Business Partners
- Independent Software Vendors

This document expands the current set of pSeries documentation by providing a desktop reference that offers a detailed technical description of the IBM @server pSeries 630 Models 6C4 and 6E4.

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