IBM @server

IBM

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

Innovative, POWER4 processor

Two unique models:
deskside/desktop or rack-mount

High-end reliability,
availability, and serviceability
features

Derrick Daines
Wolfgang Seiwald
Dave Williams

ibm.com/redbooks
sbv@us.ibm.com

International Technical Support Organization

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

August 2002

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

Second Edition (August 2002)

This edition applies to the IBM @server pSeries™ 630 Models 6C4 and 6E4 and AIX 5L™ Version 5.1, product number 5765-E61.

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 5L™	IBMLink™	Redbooks(logo)™
AIX®	Perform™	Redbooks™
Chipkill™	PowerPC®	RS/6000®
Enterprise Storage Server™	POWER4™	SP™
IBM®	pSeries™	

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

Lotus®	Word Pro®

The @server brand consists of the established IBM e-business logo followed by the descriptive term "server".

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 provides a comprehensive single-source guide covering 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

This document expands the current set of IBM @server pSeries documentation by providing a desktop reference that offers a detailed technical description about 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 Whitepaper

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

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 also provides country support during weekends.

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 area of the IBM @server systems and the IBM AIX® operating system.

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

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

IBM U.S.A.
Doug Bossen, Brad Frey, Ghadir Gholami, John Hilburn, Tenley Jackson, Jim McGaughan, Bill Mihaltse, Jan Palmer, Jim Reidy, Alan Standridge, Ven Tenjarla, and Sterling Weaver

IBM Germany
Volker Haug and Gregor Linzmeier

IBM U.K.
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 Whitepaper 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

General description

The IBM @server pSeries 630 Models 6C4 and 6E4 (referred to hereafter as the Model 6C4 and Model 6E4) are designed for customers looking for a cost-effective, high-performance 1- to 4-way server. These servers make use of IBM technology, first used in the high-end pSeries 690, delivered in a compact package. These systems use the 64-bit, copper/SOI-based, 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. They offer excellent price/performance in small packages with high-end reliability features for managing e-business, greater application flexibility, and innovative technology, all designed to help capitalize on the e-business revolution.

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 8 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 memory DIMMs.

The Models 6C4 and 6E4 include four, front-accessible disk bays supporting hot-swappable disks. These disk bays can accommodate up to 293.6 GB of disk storage using 73.4 GB Ultra3 SCSI disk drives. Two media bays are used for a CD-ROM, DVD-RAM, or another optional media device, such as a tape or diskette drive.

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 behaviors that include improved fault isolation, recovery from errors without stopping the system, avoidance of recurring failures, and predictive failure analysis. See Section 3.2, "Reliability, availability, and serviceability (RAS) features" on page 22 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] One EIA (Electronic Industries Association Unit) is 44.45mm (1.75 inch)

1.1 Minimum and optional features

The Models 6C4 and 6E4 deliver a cost-efficient growth path for the future through such capabilities as:

► 1-, 2-, and 4-way SMP system

   – 1- and 2-way processor books with 1.0 GHz POWER4 microprocessors. Only a single 1-way card is allowed per server at a time. A 3-way configuration is unavailable.

   – A 1-way (FC 5131) to a 2-way (FC 5132) upgrade is by feature conversion.

   – 32 MB shared L3 cache per processor card.

► 1 GB to 32 GB ECC DDR SDRAM memory (based on the available memory DIMMs at time of publication)

   – Memory DIMMs plug into the processor book (8 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).

   – 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 4x2048 MB).

► Four hot-swapable disk drive bays

   – 18.2 GB to 293.6 GB of disk storage.

   – 18.2 GB Ultra3 10K RPM (FC 3157), 36.4 GB Ultra3 10K RPM (FC 3158), 73.4 GB Ultra3 10K RPM (FC 3159) disks are available at the time of writing.

► Two media bays (one CD-ROM or DVD-RAM must be configured on an initial order)

   – Media bay 1 can accommodate an IDE CD-ROM (FC 2633), or a SCSI DVD-RAM (FC 2623). The DVD-RAM will read CD-ROM installation media.

   – Media bay 2 can accommodate a DVD-RAM (FC 2623), diskette drive (FC 2605), or tape drive (4 mm FC 6158 or 8 mm FC 6156; contact your IBM representative to check the availability of 8 mm tape drives).

   – FC 6203 with FC 4260 is required for the additional SCSI devices (DVD-RAM or tape drive) installed.

   – A maximum of one DVD-RAM is orderable per system.

► Four 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 14.

► Integrated ports

   – Two 10/100 Ethernet (IEEE 802.3 compliant).

   – Two Ultra3 SCSI (one external Ultra3 SCSI (with VHDCI[2] mini 68-pin port), and one internal Ultra3 SCSI disk drive backplane). VHDCI requires a mini 68-pin connector or FC 2118 mini 68-pin to 68-pin 0.3 meter cable as an additional feature.

   – 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.

   – Keyboard and mouse.

► Optional hot-plug redundant power supplies (requires the addition of redundant cooling)

► Optional redundant hot-plug cooling fans

► Integrated service processor

► AIX® 5L™ license included

► Support for 32-bit and 64-bit applications

---

[2] Very High Density Cable Interconnect (VHDCI)

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

1.2 Model types

Figure 1-1 shows the package layout for the Models 6C4 and 6E4. The Models 6E4 and 6C4 comply with the FCC limitations for a Class A digital device, and meet IBM's acoustic guidelines for a general business environment.

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

1.2.1 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 pre-configured as ordered. 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 system was designated as a customer setup and requires three persons (due to weight and safety issues) to be present to install the unit into the rack.

All system MES upgrades can be performed by the customer with the single exception of the 2-way processor card FC 5132, which requires that a service call be placed.

Each Model 6C4 is shipped with a template which helps each server be mounted into the desired position in the rack. The Model 6C4 is shipped with a cable management arm. The cable management arm provides support and routing for all cables, both power and signal, to prevent damage during movement of the system into the service position. Cables are secured to the arm using cable ties supplied.

Physical specifications

The following are the key statistics of the Model 6C4:

- ► Width: 444.4 mm (17.5 inches)
- ► Depth: 609.6 mm (24.0 inches)
- ► Height: 172.8 mm (6.8 inches)
- ► Weight in minimum configuration: 32.0 kg (70.4 pounds)
- ► Weight in maximum configuration with rails: 47.3 kg (104.0 pounds)

1.2.2 Model 6E4 deskside/desktop server

The Model 6E4 is a deskside/desktop server, 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 pre-configured as ordered. 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 server.

All system MES upgrades can be performed by the customer with the single exception of the 2-way processor card FC 5132, which requires that a service call be placed.

Physical specifications

The following are the key statistics of the Model 6E4:

- ► Width: 300.0 mm (11.8 inches)
- ► Depth: 725.0 mm (28.5 inches)
- ► Height: 530.0 mm (20.9 inches)
- ► Weight in minimum configuration: 36.0 kg (79.2 pounds)
- ► Weight in maximum configuration: 51.0 kg (112.2 pounds)

1.3 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.

1.3.1 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 classic or sculptured 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.

► Support of both AC and DC configurations.

► Up to 6 AC power distribution units (PDUs) provide a total of 36 power connections. Each PDU will require a separate AC supply. Four PDUs should be adequate for most configurations, and any subsequent PDUs require 1U of rack space.

► DC rack height is increased to 1926 mm (75.8 inches) due to the presence of the power distribution panel fixed to top of rack.

► Weight:
  – T00 base empty rack: 244 kg (535 pounds)
  – T00 full rack: 816 kg (1795 pounds)

1.3.2 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.

► Up to 7 PDUs provide a total of 42 power connections. Each PDU will require a separate AC supply. Four PDUs should be adequate for most configurations, and any subsequent PDUs require 1U of rack space.

► Weight:
  – T42 base empty rack: 261 kg (575 pounds)
  – T42 full rack: 930 kg (2045 pounds)

1.3.3 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.

1.3.4 Rack-mounting rules for Model 6C4

There are rules that should be followed when mounting drawers into a rack. The primary rules are as follows:

► 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 to ensure 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 fit in a T00 rack, or ten Model 6C4s in a T42 rack.

1.3.5 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. 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 keyboard)

Order the T210 Flat-Panel Monitor (FC 3635) if you are looking for a desk-top flat panel display option. The T210 is a 21-inch TFT LCD digital screen with a viewable image size of 528 mm (20.8 inches) measured diagonally, incorporating a 0.207 mm pixel pitch/type for bright, high-definition images.

> **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 and T210 (or any other graphics monitor) require a GXT135P Graphics Adapter (FC 2848) to be installed in the server.

1.3.6 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 and support for six languages (English, French, Spanish, German, Italian, or Brazilian Portuguese) allows fast switching between systems.

The VGA switch is only 1 EIA unit (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.

1.4 Statement of Direction

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

1.4.1 Cluster 1600

The Model 6C4, at the time of announcement, is planned to be supported as a cluster building block by the IBM @server Cluster 1600. It will be supported by PSSP 3.4 and can optionally be attached to the SP Switch2. This facility is planned for introduction in the fourth quarter of 2002.

1.4.2 LPAR

IBM is planning to enable the pSeries 630 servers with LPAR to dynamically redefine the resources that make up a logical partition. This means that partitions can be changed while applications continue to run. No re-boot of affected partitions will be required after this planned feature is installed. Processors, memory, and adapters can be logically added or removed from active partitions without disruption to applications. The LPAR requires the use of a Hardware Management Console (HMC). This capability is planned for introduction on the pSeries 630 in the fourth quarter of 2002 and will be supported by a new release of AIX 5L.

AIX 5L Version 5.1 will not support static LPAR on the pSeries 630 servers until fourth quarter of 2002.

1.4.3 Remote I/O Drawer

To enhance expandability of pSeries 630, IBM plans to introduce a 4U high rack-mount remote I/O expansion drawer with seven PCI-X slots and two six-packs of hot-swappable disk bays. This capability is expected to be introduced in the fourth quarter of 2002.

1.4.4 NEBS Level 3 and -48 Volts DC

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

---

[3] Network Equipment Building System

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 using production workloads.
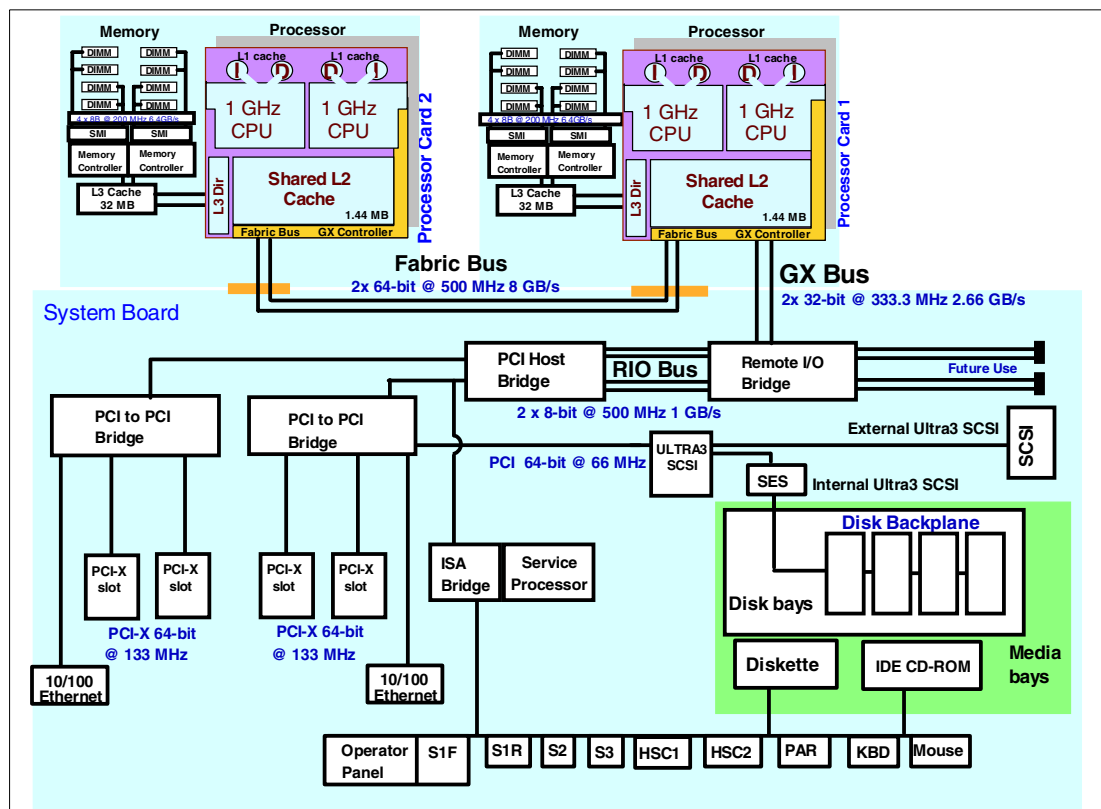

Figure 2-1   Conceptual diagram of the Models 6C4 and 6E4 system architecture

2.1 Processor and cache

The Models 6C4 and 6E4 processor subsystem consists of POWER4 processors running at 1.0 GHz. These are the same processor chips used in the IBM @server pSeries 690. However, unlike the Multi-Chip 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 to 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 (which was used between chips on the same MCM in the Model 690) is no longer relevant in the SCM. A logical diagram of the pSeries 630 SCM and the pSeries 690 MCM is shown in Figure 2-2 for comparison.
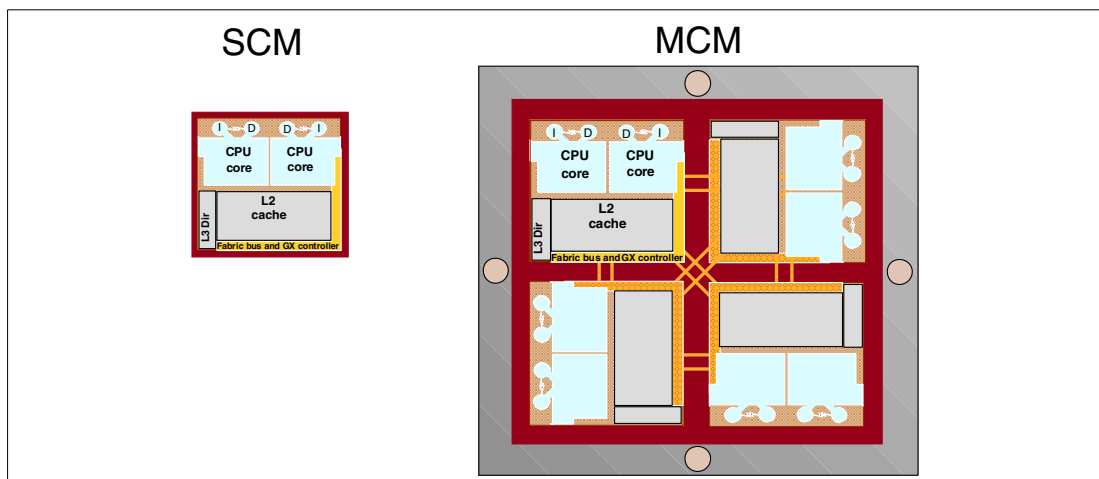
Figure 2-2   Comparison between SCM and MCM

Each SCM is a Ceramic Column Grid Array (CCGA) package where the chip carrier is raised slightly from it's 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 named a *book* that protects (both in and out of the server) and secures the card, and helps manage airflow used for cooling.
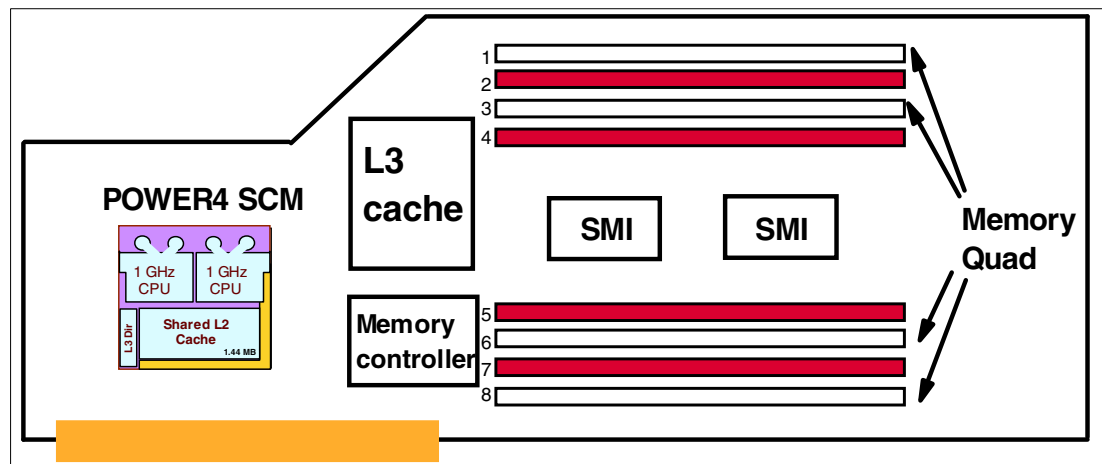
Figure 2-3   Processor card layout

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 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 Dat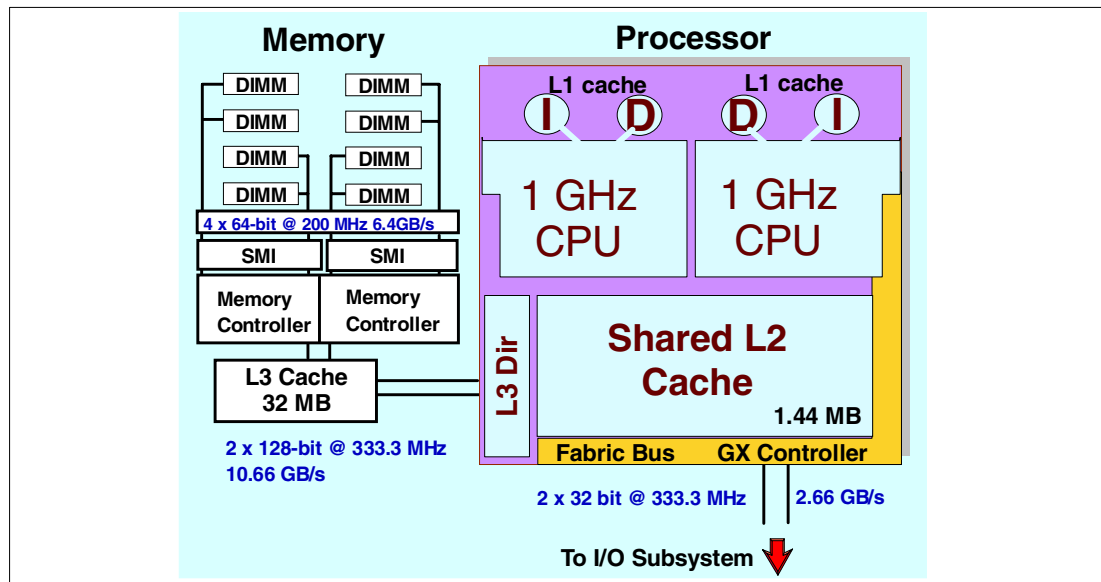a (D) cache per processor core. The second level is 1.44 MB of L2 cache. A 2-way processor shares the L2 cache. All cache has either full ECC[1] or parity protection on the data arrays, and L1 cache has the ability to re-fetch data from L2 cache in the event of soft errors detected by parity checking.

The Level 3 (L3) cache consists of two components: the L3 cache controller/directory and the L3 data array. The L3 controller and cache directory are on the POWER4 chip, and the data array itself consists of two 16 MB embedded DRAM (eDRAM) chips 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 0.18-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 lower power consumption and improved reliability.

---

[1] ECC single error correct, double error detect
[2] Complementary Metal Oxide Semiconductor

2.1.4 Processor clock rate

The Models 6C4 and 6E4 operate with a processor clock rate of 1.0 GHz.

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

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

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

`pmcycles -m`   This command uses the Performance Monitor cycle counter and the processor real-time clock to measure the actual processor clock speed in 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 11. As shown, there are four 64-bit data paths to 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 (see Figure 2-4 on page 11).

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 (4-way) for best performance, therefore a 1 GB 4-way configuration is not recommended.

2.2.1 Memory options

The following memory features for the Models 6C4 and 6E4 are available at the time of writing:

FC 4451   1024 MB (4X256 MB) 208 pin 8 ns DDR SDRAM DIMMs

FC 4452   2048 MB (4X512 MB) 208 pin 8 ns DDR SDRAM DIMMs

FC 4453   4096 MB (4X1024 MB) 208 pin 8 ns stacked DDR SDRAM DIMMs

FC 4454   8192 MB (4X2048 MB) 208 pin 8 ns stacked DDR SDRAM DIMMs

Each memory feature consists of four DIMMs, or quad.

---

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

2.2.2 Memory guidelines for LPAR

Planning for logical partitioning (see Section 1.4, "Statement of Direction" on page 7) involves the following considerations regarding memory.

► pSeries 630s must have at least 1 GB to support LPAR. The first 256 MB of memory will be allocated to the hypervisor, the last 256 MB to translation control entries (TCEs) and to hypervisor per partition page tables. This minimum configuration will allow one 512 MB partition or two 256 MB partitions (given a hypervisor memory allocation unit of 256 MB). TCE memory is used to translate the I/O addresses to system memory addresses. Without LPAR, this memory is part of the memory AIX allocates. In LPAR mode, where a system can have many AIX LPAR images, the TCE memory is allocated once for the set of images rather than once for each LPAR image.

► The memory installed on processor card 1 should match the size of the memory installed on processor card 2.

► With AIX 5L Version 5.1, pSeries 630s can never support a partition with a memory size 16 GB or greater even if more than 16 GB of physical memory is installed.

2.2.3 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 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.
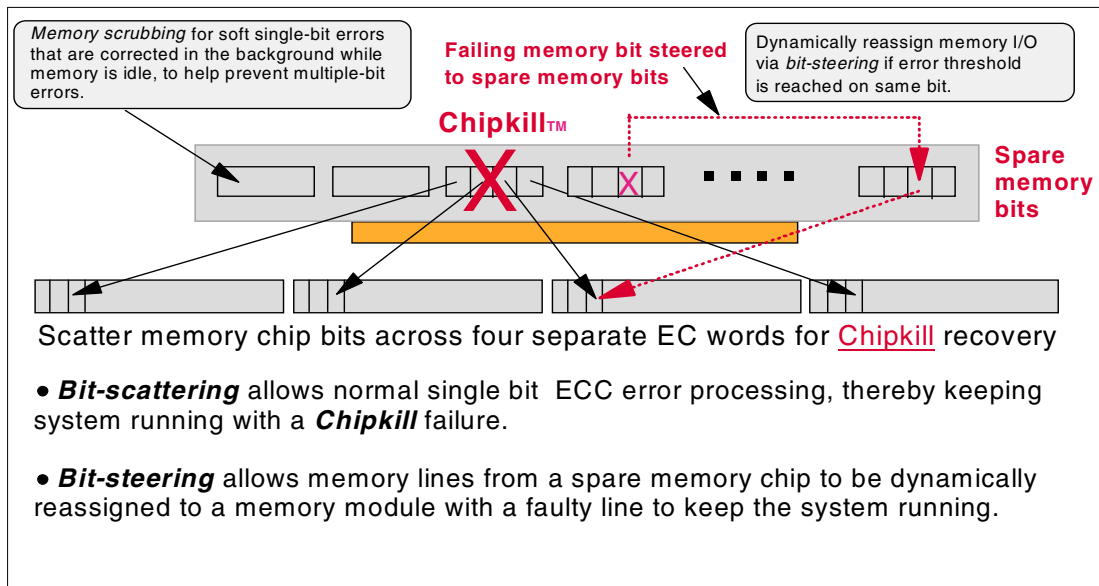

Figure 2-5   Main storage ECC and extensions

2.3 System buses

The system bus from the processors to the memory subsystem is 2 x 128 bits at 1/3 CPU clock speed (or 333.3 MHz) for an aggregate data rate of 10.66 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 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 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 32-bit paths at 333.3 MHz to give an aggregate data rate of 2.66 GB/s.

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 8 bits at 500 MHz in each direction, or 1 GB/s bidirectionally.

The second processor card has access to the GX bus using a Fabric bus, which connects it to the first processor card. The Fabric bus is similar in nature to the GX bus, but has dual 64-bit paths operating at 500 MHz to provide an aggregate data rate of 8 GB/s.

2.3.2 PCI host bridge and PCI bus

Two RIO ports connect to the PCI Host Bridge (PHB). The PHB chip acts as a bridge between the RIO bus and two PCI to PCI bridges, which fan out to the separate PCI buses.

In the current implementation, the PCI to PCI bridge supports four PCI-X slots and two Ultra3 SCSI controllers on the system board, which 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 the system reboot.

In general, 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

In general, PCI-X slots also support existing 3.3 volt PCI adapters, which include 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.

2.4.2 Internal Ultra3 SCSI controllers

The internal 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.

For the attachment of any other internal devices (for example CD-ROM, DVD-RAM, or tape), an additional SCSI card and appropriate cables are required.

2.4.3 LAN adapters

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

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

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

2.4.4 Graphics adapters

The Models 6C4 and 6E4 are servers and are not intended to serve as a workstation. Therefore, the GXT135P is the only graphics adapter available. This adapter offers 2D function for business graphics, Internet applications, or for those applications that require a graphics display for installation and management.

2.5 Internal storage

The internal storage on the Models 6C4 and 6E4 are similar in all respects and so will be discussed together.

The storage contained within the Models 6C4 and 6E4 consists of both SCSI and IDE devices, with the IDE devices being CD-ROM drives and the SCSI devices being internal disks, tape, and DVD-RAM, some of which will require a SCSI adapter to be placed in one of the PCI slots, as discussed in the next section.

---

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

2.5.1 Internal media

The Models 6C4 and 6E4 have two internal media bays. These bays can be populated by the devices listed in Table 2-1.

Table 2-1   List of internal media options

Feature Code	Description	Comments
2633	650 MB IDE 48x CD-ROM Drive	This drive requires cable FC 4261 to connect the drive to the system board.
2605	Diskette Drive	This diskette drive, when fitted, occupies a full media bay.
2623	DVD-RAM Drive 4.7 GB capacity	This feature is a SCSI device and will require one of the following additional feature codes: 6203 with 4260. Either one of these adapters will necessitate the use of a PCI slot within the machine.
6158	4 mm 20/40 GB Tape Drive	This feature is a SCSI device and will require one of the following additional feature codes: 6203 with 4260. Either one of these adapters 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) is sufficient for supporting SCSI devices in both the internal media bays.		

2.5.2 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 services, 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 2-2.

Table 2-2   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
3159	73.4 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.

2.5.3 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 on-line 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-swap 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.

2.5.4 Boot options

Both Models 6C4 and 6E4 handle the boot process in a similar way to other RS/6000 and 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 or 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 a `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 above. For more information on 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 FC 6228 with microcode 3.22A1 or later. 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

2.5.5 I/O guidelines for LPAR

Planning for logical partitioning (see Section 1.4.2, "LPAR" on page 7) requires the following considerations with respect to I/O:

- ► The following I/O must be assigned to the same partition:
    - PCI slot 1 and slot 2
    - Integrated SCSI ports
    - Ethernet port *XXX*
    - All hot-swap internal disks
    - IDE CDROM
    - Diskette drive
- ► PCI slot 3 and slot 4 and Ethernet port *YYY* can be assigned to other partitions.
- ► An HMC (such as from a p690) must be available to manage LPAR configurations.
- ► Additional partitions must have an installation method, such as NIM and means of running diagnostics, such as network diagnostics.
- ► Additional partitions will require dedicated access to a boot device located in an I/O drawer or external enclosure, as well as devices used for all other non-shared data such as user data or system paging.

2.6 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.

2.7 Operating system requirements

The Models 6C4 and 6E4 are capable of running IBM AIX 5L for POWER and are planned to support appropriate versions of Linux when available. AIX 5L has been specifically developed and enhanced to exploit and support the advanced RAS features on IBM @server pSeries systems.

2.7.1 AIX 5L

The Models 6C4 and 6E4 require AIX 5L Version 5.1 at Maintenance Level 2 or later, plus APAR IY31315. In order to boot from CD, make sure you have the following media:

- ► AIX 5L Version 5.1 5765-E61, dated 04/2002 (CD# LCD4-1061-03) or later.

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 5.1 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.

AIX 5L 32-bit versus 64-bit kernel

On the 64-bit hardware, which the Models 6C4 and 6E4 provide, you can run either a 32-bit or 64-bit AIX kernel. During the installation process, both kernel filesets are installed. For information on how to switch between them without reinstallation, see the *64-bit Mode on AIX Version 4.3.3 to 5.1.0* paper in the *Tips for AIX Administrators* database, found on the Internet at:

http://techsupport.services.ibm.com/server/aix.techTips

Both the 32-bit and 64-bit kernels provide similar application support through common Application Binary Interfaces (ABIs), so the vast majority of applications will run on both kernel types. You can even run 32-bit and 64-bit applications concurrently on the same system. But be aware that a 64-bit kernel requires that all kernel extensions and device drivers are 64-bit, while a 32-bit kernel requires them to be 32-bit. Make sure all in-house and third party kernel extensions and device drivers are ported accordingly.

2.7.2 Linux

A Linux® distribution, at the time of writing, is available through SuSE, for an overview of this support, please see:

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

Full information on SuSE Linux Enterprise Server 7 for pSeries, see:

http://www.suse.com/us/products/suse_business/sles/sles_iSeries_pSeries/index.html

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

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

Availability, investment protection, expansion, and accessibility

The following sections provide more detailed information about configurations, upgrades, and design features that will help you lower your cost of ownership.

3.1 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 have logically 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. Only for such demand, the S2 port also can be used for HACMP. When the machine is at 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 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 the port is used by HACMP, and if Ctrl+D code is coming in from S2 port, then service processor disables the service processor menu for S2 port. This menu disablement lasts until the next time the service processor is reset, either by 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 S3 port for UPS and S2 port for HACMP. Also, once you have set up this configuration decide 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 port 2 and 3 supports HACMP heartbeat functions. If the S2 is used for HACMP heartbeat then it can't be used again for 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.

3.2 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 operate when required, to perform reliably, and to efficiently handle any failures which may occur. This is achieved using capabilities provided by both the hardware and the AIX 5L operating system.

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 re-create 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 for power, cooling, and I/O subsystems.

► Predictive failure analysis on processors, caches, memory, I/O, 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 (processor boot-time deallocation based on run-time errors).

► Persistent deallocation extended to caches and 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 IBM Model 690 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 RAS features, such as Chipkill memory, are explained in previous sections of this document (see Section 2.2, "Memory" on page 12). Additional RAS features of the Models 6C4 and 6E4 are covered in the following sections.

3.2.1 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 Section 3.2.4, "Service processor" on page 25) and in run-time First Failure Data Capture (FFDC) 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.

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 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.

3.2.2 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 subsequent boot time.

Dynamic deallocation functions include:

- ► Processor and L2 cache

- ► L3 cache line delete

- ► PCI bus

For Dynamic Processor Deallocation, the service processor performs a predictive failure analysis based on any recoverable processor errors which have been recorded. If these transient errors exceed a defined threshold, the event is logged and the processor is

de-allocated from the system while the operating system continues to run. This feature (named *cpuguard*) enables maintenance to be deferred to 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:** With AIX 5L Version 5.1, the use of cpuguard is only effective on systems with three or more functional processors.

Cache or cache-line deallocation is aimed at performing dynamic re-configuration to bypass potentially failing components. This capability is provided for both L2 and L3 cache. 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 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-swap 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 shutdown. 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
► L2 and L3 cache
► Deconfigure or bypass failing I/O adapters

Following a hardware error which has been flagged by the service processor, the subsequent re-boot 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).

3.2.3 Power and cooling

Environmental monitoring related to power, fan operation, and temperature is done by the System Power Control Network (SPCN). 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 the minimum of disruption. Note that when ambient temperatures exceed 32C (92F), 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 either with 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 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.

3.2.4 Service processor

The service processor (SP) is a specialized device that is situated on the system board and provides a number of different functions, discussed in the following:

- ► 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 SP, 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 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.

3.2.5 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 in the following:

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 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. 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. A hardware fault will also turn on the two attention indicators (one located on the front and the other located on the rear of the system) to alert the user of a hardware problem.

3.2.6 Project eLiza

IBM's project eLiza is about using technology to manage technology. Project eLiza 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.

eLiza on IBM @server pSeries servers[6] describes the many self-configuring, self-optimizing, self-protecting, and self-healing features which are available on IBM @server pSeries servers.

Self configuring　Project eLiza's strategy is to provide 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.

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.

---

[6] http://www-1.ibm.com/servers/eserver/introducing/eliza/pseries.html

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 which are as secure as the original resource management systems.

Self optimization Self-optimization requires a system to efficiently maximize resource utilization to meet the end user needs with no human intervention required. Today's systems already include industry leading technologies like logical partitioning, dynamic workload management and dynamic server clustering.

Recently announced capabilities included items like: LPAR and dynamic LPAR (planned for forth quarter 2002), workload manager, Reliable Scalable Cluster Technology (RSCT), and a self-protecting kernel. Kerberos V5 authentication, IP-address takeover and Virtual IP Address (VIPA). These features depend upon the inherent characteristics of the hardware and AIX 5L operating system.

Related publications

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

IBM Redbooks

For information on ordering these publications, see "How to get IBM Redbooks" on page 30.

- *Understanding IBM @server pSeries Performance and Sizing,* SG24-4810
- *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

Other resources

These publications are also relevant as further information sources:

- *7014 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.
- *7316-TF2 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 User's Guide*, SA38-0606, contains information to help users use the system, use the service aids, and solve minor problems.
- *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.
- *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 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:

► The POWER4 system microarchitecture is 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

► CCGA - Ceramic Column Grid Array, see IBM Chip Packaging

http://www.ibm.com/chips/micronews

► SOI - Silicon on Insulator technology

http://www-3.ibm.com/chips/bluelogic/showcase/soi

► Copper circuitry

http://www-3.ibm.com/chips/bluelogic/showcase/copper/

► CHRP - PowerPC Microprocessor Common Hardware Reference Platform: A System Architecture

http://www.mkp.com/books_catalog/catalog.asp?ISBN=1-55860-394-8

► Hardware documentation

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

► Frequently asked SSA related questions

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

► 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

► 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/

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.

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

Innovative, POWER4 processor

Two unique models: deskside/desktop or rack-mount

High-end reliability, availability, and serviceability features

This document provides a comprehensive single-source 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

This document expands the current set of pSeries documentation by providing a desktop reference that offers a detailed technical description about 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.