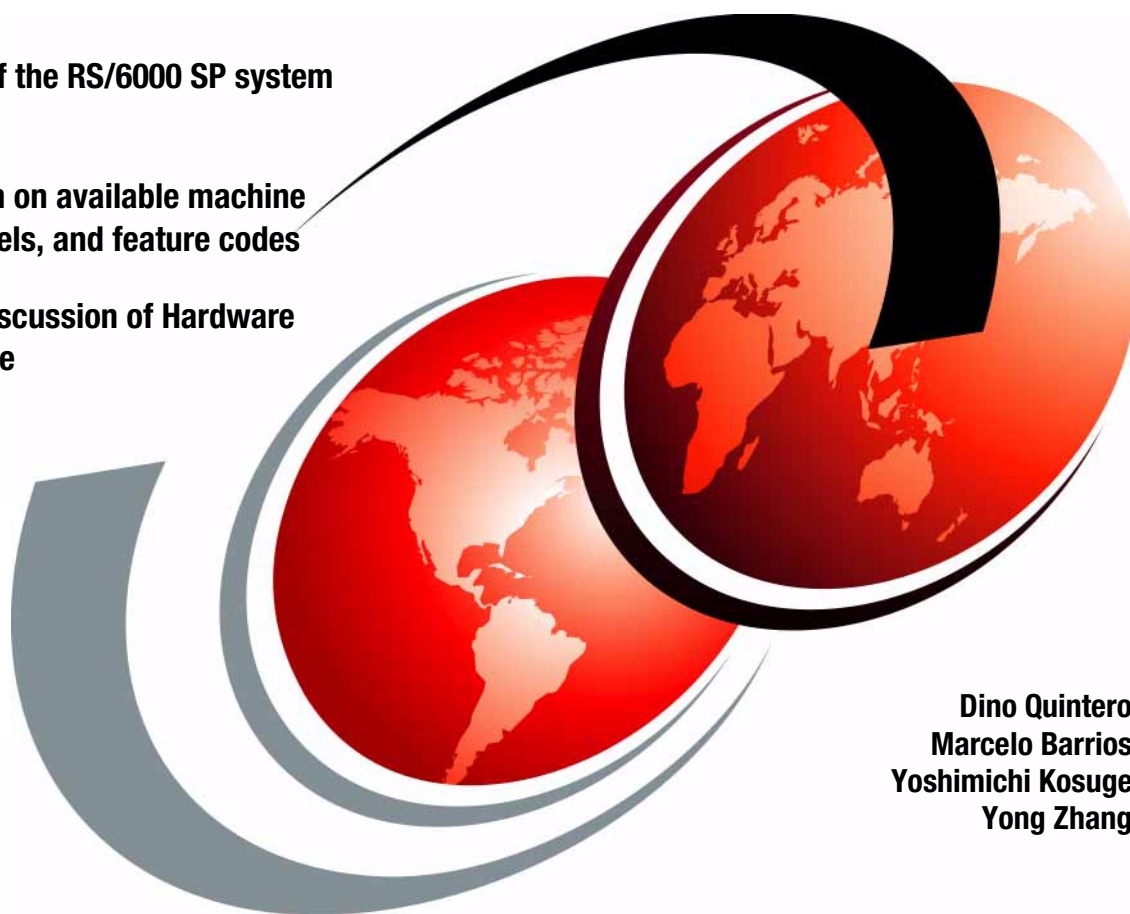


RS/6000 SP Systems Handbook

Overview of the RS/6000 SP system

Information on available machine types, models, and feature codes

In-depth discussion of Hardware Architecture



Dino Quintero
Marcelo Barrios
Yoshimichi Kosuge
Yong Zhang

ibm.com/redbooks

Redbooks



International Technical Support Organization

RS/6000 SP Systems Handbook

October 2000

Take Note!

Before using this information and the product it supports, be sure to read the general information in Appendix A, "Special notices" on page 205.

Second Edition (October 2000)

This edition applies to IBM RS/6000 SP. Related software offerings include IBM Parallel System Support Programs for AIX Version 3, Release 2 (5765-D51) and AIX Version 4, Release 3.

Comments may be addressed to:
IBM Corporation, International Technical Support Organization
Dept. JN9B Mail Station P099
2455 South Road
Poughkeepsie, NY 12601-5400

When you send information to IBM, you grant IBM a non-exclusive right to use or distribute the information in any way it believes appropriate without incurring any obligation to you.

© Copyright International Business Machines Corporation 2000. All rights reserved.
Note to U.S Government Users – Documentation related to restricted rights – Use, duplication or disclosure is subject to restrictions set forth in GSA ADP Schedule Contract with IBM Corp.

Contents

Figuresix
Tablesxi
Prefacexiii
The team that wrote this redbookxiii
Comments welcomexiv
Chapter 1. Introduction	1
1.1 Overview	1
1.1.1 Origins	2
1.1.2 Management and availability	4
1.1.3 Features and benefits	5
1.1.4 Business solutions	7
1.2 Hardware components	9
1.2.1 Frames	11
1.2.2 Processor nodes	11
1.2.3 Extension nodes	13
1.2.4 Switches	14
1.2.5 Control workstations	14
Chapter 2. Frames	15
2.1 Short frames	17
2.1.1 Short model frames (Model 500)	18
2.1.2 Short expansion frames (F/C 1500)	19
2.2 Tall frames	20
2.2.1 Tall model frames (Model 550)	22
2.2.2 Tall expansion frames (F/C 1550)	24
2.3 Switch frames	26
Chapter 3. Processor nodes	29
3.1 POWER3-II SMP nodes	30
3.1.1 POWER3-II microprocessor	31
3.2 POWER3 SMP High Node (F/C 2058)	35
3.2.1 PCI bus description	36
3.2.2 Requirements	36
3.2.3 Options	36
3.2.4 Processor requirements and options	37
3.2.5 Disk requirements and options	38
3.2.6 Switch and communication adapter requirements and options ..	38
3.2.7 375 MHz POWER3 SMP Wide Node (F/C 2057)	39

3.2.8	375 MHz POWER3 SMP Thin Node (F/C 2056)	44
3.3	POWER3 SMP nodes	48
3.3.1	POWER3 SMP node system architecture	48
3.3.2	POWER3 SMP High Node (F/C 2054)	52
3.3.3	PCI bus description	52
3.3.4	Requirements	53
3.3.5	Options	53
3.3.6	Processor requirements and options	53
3.3.7	Disk requirements and options	54
3.3.8	Switch and communication adapter requirements and options	55
3.3.9	SP Expansion I/O unit (F/C 2055)	56
3.3.10	DASD options	56
3.4	332 MHz SMP nodes	57
3.4.1	332 MHz SMP node system architecture	58
3.4.2	332 MHz SMP thin nodes (F/C 2050)	64
3.4.3	332 MHz SMP wide nodes (F/C 2051)	67
Chapter 4. SP-attached servers		73
4.1	Overview	73
4.1.1	How the SP system views the SP-attached server	74
4.2	Installation requirements	76
4.2.1	System requirements	76
4.2.2	RS/6000 SP System Attachment adapter (RS/6000 F/C 8396)	77
4.2.3	Network media card requirements	79
4.2.4	Software requirements	81
4.3	Network interface	83
4.3.1	Connecting to the control workstation	86
4.3.2	Connecting to the SP Switch	88
Chapter 5. Clustered Enterprise Server System		91
5.0.1	Clustered Enterprise Server overview	91
5.0.2	Clustered Enterprise Server installation requirements	91
5.0.3	Clustered Enterprise Server system requirements	91
5.0.4	Clustered Enterprise Server network media card requirements	92
5.0.5	Clustered Enterprise Server software requirements	92
5.0.6	Planning the Clustered Enterprise Server network	93
Chapter 6. SP Switch Routers		97
6.1	Overview	97
6.2	Installation requirements	101
6.2.1	System requirements	102
6.2.2	SP Switch Router adapter (F/C 4021)	102
6.2.3	Network media card requirements	103
6.2.4	Software requirements	104

6.3 Network interface	104
6.3.1 Connecting to the control workstation	105
6.3.2 Connecting to the SP Switch	107
Chapter 7. SP Switch network	109
7.1 Overview	109
7.1.1 SP Switch board	110
7.1.2 SP Switch adapter	112
7.1.3 SP Switch network	114
7.2 SP Switch products	115
7.2.1 SP Switch2 (F/C 4012)	116
7.2.2 SP Switch (F/C 4011)	116
7.2.3 SP Switch-8 (F/C 4008)	117
7.2.4 SP Switch2 adapter (F/C 4025)	118
7.2.5 SP Switch adapter (F/C 4020, 4022, and 4023)	120
7.2.6 Special SP Switch adapter (F/C 4021 and F/C 8396)	121
Chapter 8. Control workstations	123
8.1 Overview	123
8.2 Installation requirements	124
8.2.1 Supported RS/6000 workstations	125
8.2.2 System requirements	126
8.2.3 Interface adapter requirements	128
8.2.4 Software requirements	130
8.3 High Availability Control Workstation	131
8.3.1 Overview	131
8.3.2 Installation requirements	135
Chapter 9. Communication adapters	137
9.1 PCI nodes communication adapters	137
9.1.1 FDDI SK-NET LP SAS (F/C 2741)	138
9.1.2 FDDI SK-NET LP DAS (F/C 2742)	139
9.1.3 FDDI SK-NET UP DAS (F/C 2743)	141
9.1.4 S/390 ESCON Channel adapter (F/C 2751)	142
9.1.5 Token Ring Auto LANstreamer (F/C 2920)	143
9.1.6 EIA 232/RS-422 8-Port Asynchronous adapter (F/C 2943)	144
9.1.7 WAN RS232 128-port (F/C 2944)	146
9.1.8 ARTIC960Hx 4-port Selectable adapter (F/C 2947)	148
9.1.9 2-port Multiprotocol X.25 adapter (F/C 2962)	149
9.1.10 ATM 155 TURBOWAYS UTP adapter (F/C 2963)	151
9.1.11 Ethernet 10/100 MB (F/C 2968)	153
9.1.12 Gigabit Ethernet - SX adapter (F/C 2969)	154
9.1.13 Ethernet 10 MB BNC (F/C 2985)	157
9.1.14 Ethernet 10 MB AUI (F/C 2987)	158

9.1.15	ATM 155 MMF (F/C 2988)	159
9.1.16	Ultra SCSI Single Ended adapter (F/C 6206)	160
9.1.17	Ultra SCSI Differential (F/C 6207)	162
9.1.18	SCSI-2 F/W Single-Ended adapter (F/C 6208)	163
9.1.19	SCSI-2 F/W Differential adapter (F/C 6209)	165
9.1.20	SSA RAID 5 adapter (F/C 6215)	166
9.1.21	SSA Fast-Write Cache Module (F/C 6222)	166
9.1.22	ARTIC960RxD Quad Digital Trunk adapter (F/C 6310)	167
9.1.23	IBM Network Terminal Accelerator (256 Session - F/C 2402)	168
9.1.24	IBM Network Terminal Accelerator (2048 Session - F/C 2403)	169
9.1.25	SCSI-2 High Performance External I/O Controller (F/C 2410)	169
9.1.26	Enhanced SCSI-2 Differential Fast/Wide adapter/A (F/C 2412)	169
9.1.27	SCSI-2 Fast/Wide adapter/A (F/C 2415)	170
9.1.28	SCSI-2 Differential Fast/Wide adapter/A (F/C 2416)	170
9.1.29	SCSI-2 Differential External I/O Controller (F/C 2420)	170
9.1.30	4-Port Multiprotocol Communications Controller (F/C 2700)	170
9.1.31	FDDI attachment (F/C 2724 and F/C 2723)	171
9.1.32	HIPPI (F/C 2735)	171
9.1.33	S/390 ESCON Channel Emulator adapter (F/C 2754)	171
9.1.34	Block Multiplexer Channel adapter - BMCA (F/C 2755)	172
9.1.35	ESCON Control Unit adapter (F/C 2756)	172
9.1.36	8-Port Async adapter - EIA-232 (F/C 2930)	172
9.1.37	8-Port Async adapter - EIA-422A (F/C 2940)	172
9.1.38	X.25 Interface Co-Processor/2 (F/C 2960)	172
9.1.39	Token Ring High Performance Network adapter (F/C 2970)	172
9.1.40	Auto Token-Ring LANstreamer MC 32 adapter (F/C 2972)	173
9.1.41	Ethernet High Performance LAN adapter (F/C 2980)	173
9.1.42	TURBOWAYS 100 ATM adapter (F/C 2984)	173
9.1.43	TURBOWAYS 155 ATM adapter (F/C 2989)	173
9.1.44	Ethernet LAN adapter (AUI/10BaseT) (F/C 2992)	173
9.1.45	Ethernet LAN adapter 10Base2 (BNC) (F/C 2993)	174
9.1.46	10/100 Ethernet Twisted Pair MC adapter (F/C 2994)	174
9.1.47	Ethernet 10BaseT Transceiver (F/C 4224)	174
9.1.48	9333 High Performance Subsystem adapter (F/C 6212)	174
9.1.49	SSA 4-Port adapter (F/C 6214)	174
9.1.50	Enhanced SSA 4-Port adapter (F/C 6216)	174
9.1.51	SSA 4-Port RAID adapter (F/C 6217)	175
9.1.52	SSA Multi-Initiator/RAID EL adapter (F/C 6219)	175
9.1.53	SSA Fast-Write Cache Module (F/C 6222)	175
9.2	SP-attached servers communication adapters	175
9.3	SP Switch Routers network media cards	176
9.3.1	ATM OC3, two port SM fiber (F/C 1101)	177
9.3.2	ATM OC3, two port MM fiber (F/C 1102)	177

9.3.3	SONET/IP OC3, one port MM fiber (F/C 1103)	177
9.3.4	SONET/IP OC3, one port SM fiber (F/C 1104)	178
9.3.5	ATM OC12, one port SM fiber (F/C 1105)	178
9.3.6	FDDI, four port MM fiber (F/C 1106)	178
9.3.7	Ethernet 10/100Base-T, eight port (F/C 1107)	178
9.3.8	HIPPI, one port (F/C 1108)	178
9.3.9	HSSI, two port (F/C 1109)	178
9.3.10	Ethernet 10/100Base-T, four port (F/C 1112)	178
9.3.11	Blank faceplate (F/C 1113)	178
9.3.12	64 MB DRAM SIMM (F/C 1114)	179
9.3.13	ATM OC12, one port MM fiber (F/C 1115)	179
9.3.14	SP Switch Router adapter (F/C 4021)	179
9.3.15	SP Switch Router adapter cable - 10 Meter (F/C 9310)	179
9.3.16	SP Switch Router adapter cable - 20 Meter (F/C 9320)	179
Chapter 10. Software support		181
10.1	Parallel System Support Programs (PSSP)	181
10.1.1	Advantages	181
10.1.2	Description	181
10.2	General Parallel File System (GPFS)	186
10.2.1	Advantages	186
10.2.2	Description	187
10.3	LoadLeveler	189
10.3.1	Advantages	189
10.3.2	Description	189
10.4	Parallel Engineering and Scientific Subroutine Library (PESSL)	192
10.4.1	Advantages	192
10.4.2	Description	192
10.5	Parallel Optimization Subroutine Library (OSLp)	195
10.5.1	Advantages	195
10.5.2	Description	196
10.6	Parallel Environment (PE)	199
10.6.1	Description	199
10.7	Performance Toolbox Parallel Extensions (PTPE)	202
10.7.1	Advantages	202
10.7.2	Description	203
Appendix A. Special notices		205
Appendix B. Related publications		209
B.1	IBM Redbooks	209
B.2	IBM Redbooks collections	209
B.3	Other resources	210
B.4	Referenced Web sites	210

How to get IBM Redbooks	211
IBM Redbooks fax order form	212
List of Abbreviations	213
Index	217
IBM Redbooks review	225

Figures

1. RS/6000 SP system sample configuration	10
2. SP frame (model 550)	16
3. Front view of short frame	18
4. Front and rear views of tall frame	22
5. Front view of SP Switch frame with eight Intermediate switch boards.	27
6. SP nodes (332 MHz thin node)	29
7. POWER3-II SMP node architecture.	31
8. POWER3-II SMP Thin node package	34
9. POWER3-II SMP Wide node package.	35
10. POWER3 SMP node system architecture block diagram	49
11. 332 MHz SMP node system architecture block diagram	59
12. IBM RS/6000 7017 Enterprise Server	73
13. SP-attached server network connections	85
14. IBM 9077 SP Switch Router model 04S (left) and 16S (right)	97
15. SP Switch Router configuration	99
16. Rear view of SP Switch Router model 04S	100
17. Rear view of SP Switch Router model 16S	101
18. SP Switch Router network interface	105
19. Front view of SP Switch board	109
20. SP Switch board	110
21. SP Switch chip link and SP Switch chip port	111
22. SP Switch chip	112
23. SP Switch adapter	113
24. Internal bus architecture for PCI nodes	114
25. SP Switch network	115
26. SP Switch MX Adapter with 332 MHz SMP Node	119
27. SP Switch2 Adapter Hardware Structure	119
28. Control workstation interface	124
29. High Availability Control Workstation with disk mirroring	132

Tables

1.	375 MHz POWER3 SMP High Node (F/C 2058) processor options	37
2.	375 MHz POWER3 SMP High Node (F/C 2058) memory features	37
3.	375 MHz POWER3 SMP Wide Node (F/C 2057) processor options	41
4.	375 MHz POWER3 SMP Wide Node (F/C 2057) memory features	41
5.	375 MHz POWER3 SMP Thin Node (F/C 2056) processor options	45
6.	375 MHz POWER3 SMP Thin Node (F/C 2056) memory features	46
7.	POWER3 SMP High Node (F/C 2054) processor options	54
8.	POWER3 SMP High Node (F/C 2054) memory features	54
9.	332 MHz SMP Thin Node processor options	66
10.	332 MHz SMP Thin Node memory features	66
11.	332 MHz SMP Wide Node processor options	70
12.	332 MHz SMP Wide Node memory features	71
13.	Supported communication adapters for SP-attached servers	80
14.	SP Switch Router network media cards and other options	103
15.	SP Switch adapter features	121
16.	Special SP Switch adapter features	122
17.	Supported RS/6000 PCI control workstations	125
18.	Supported RS/6000 MCA control workstations	126
19.	Serial port adapters for PCI control workstations	128
20.	Serial port adapters for MCA control workstations	129
21.	Ethernet adapters for PCI control workstations	130
22.	Ethernet adapters for MCA control workstations	130
23.	Supported communication adapters for PCI nodes	137
24.	Cable information for 2-port Multiprotocol adapter	151
25.	Supported communication adapters for SP-attached servers	175
26.	SP Switch Router network media cards	176

Preface

This redbook is a comprehensive guide dedicated to the RS/6000 SP product line. Major hardware and software offerings are introduced and their prominent functions are discussed.

This publication is suitable for the following professionals who wish to acquire a better understanding of RS/6000 SP products, including:

- Customers
- Sales and marketing professionals
- Technical support professionals
- Business partners

Inside this publication, you will find:

- An overview of the RS/6000 SP system
- Discussion of hardware architecture
- Information on available hardware machine types, models, and feature codes
- Information on supported communication adapters for the RS/6000 SP system
- Information on software available for the RS/6000 SP system

This redbook does not replace the latest RS/6000 SP 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 solutions for the UNIX marketplace using RS/6000 SP systems.

The team that wrote this redbook

This redbook was produced by a team of specialists from around the world working at the Poughkeepsie Center.

Quintero Dino is a project leader at the International Technical Support Organization (ITSO), Poughkeepsie Center. He has over nine years of experience in the Information Technology Field. He holds a BS in Computer Science, and a MS degree in Computer Science from Marist College. He has been with IBM since 1996. His areas of expertise include Enterprise Backup and Recovery, Disaster Recovery Planning and Implementation, and RS/6000. He is also a Microsoft Certified Systems Engineer. Currently, he

focuses on RS/6000 Cluster Technology by writing redbooks and teaching classes worldwide.

Marcelo Barrios is a project leader at the International Technical Support Organization, Poughkeepsie Center. He has been with IBM since 1993 and has worked in different areas related to RS/6000. Currently, he focuses on RS/6000 SP technology by writing redbooks and teaching IBM classes worldwide.

Yoshimichi Kosuge is an IBM RS/6000 SP project leader at the International Technical Support Organization, Poughkeepsie Center. Since he joined IBM Japan, he has worked in the following areas: LSI design, S/390 CP microcode, VM, MVS, OS/2, and AIX. After joining the ITSO in 1998, he has been involved in writing redbooks and teaching IBM classes worldwide on all areas of RS/6000 SP.

Yong Zhang is an Advisory Technical Specialist in Beijing, China. He has worked at IBM China for about five years. He is responsible for providing technical and marketing support for the RS/6000 product family.

Thanks to the following people for their invaluable contributions to the first edition of this book:

IBM Poughkeepsie

Joseph Banas, Endy Chiakpo, Joan McComb, Linda Mellor, Mary Nisley, John Simpson, William Tuel

IBM Somers

Barbara M Butler

Comments welcome

Your comments are important to us!

We want our Redbooks to be as helpful as possible. Please send us your comments about this or other Redbooks in one of the following ways:

- Fax the evaluation form found in “IBM Redbooks review” on page 225 to the fax number shown on the form.
- Use the online evaluation form found at ibm.com/redbooks
- Send your comments in an Internet note to redbook@us.ibm.com

Chapter 1. Introduction

The RS/6000 SP high-performance system uses the power of parallel processing to expand your application horizons. Designed for performance and scalability, this system makes feasible the processing of applications characterized by large scale data handling and compute intensity.

1.1 Overview

The SP system simultaneously brings dozens to hundreds of RISC processor nodes to a computing problem. Its parallel processing capability enhances computing performance and throughput many times over serial computing. In addition to helping improve the performance of existing applications, new applications, like complex data mining and modeling of the universe, are now possible.

The basic SP building block is the processor node. It consists of a POWER3 symmetric multiprocessor (SMP), PowerPC SMP, or POWER2 Super Chip (P2SC) uniprocessor, memory, PCI or Micro Channel expansion slots for I/O and connectivity, and disk devices. The three sizes of nodes (thin, wide, and high) may be mixed in a system and are housed in short or tall system frames. Depending on the sizes of the nodes used, an SP tall frame can contain up to 16 nodes. These frames can be interconnected to form a system with up to 128 nodes (512 by special order) where a maximum of 64 SMP high nodes can be installed per system.

The POWER3 SMP nodes are powered by the same POWER3 processor technology introduced with the 43P model 260 workstation. Hence, all applications that run on the 43P model 260 should run unchanged on a single node of the SP system. As an example, Environmental Systems Research Institute, without having tested their applications specifically on the POWER3 SMP node, has stated that their applications ARC/INFO, ArcView, SDE, and IMS will work fine on this SP node.

The 332 MHz SMP nodes are powered by the PowerPC 604e processor. They represent the first general availability of the advanced technology used in the SP system IBM has delivered to Lawrence Livermore National Laboratory as part of the Department of Energy's Accelerated Strategic Computing Initiative (ASCI) project. This system will perform the complex calculations required for the simulation to predict the performance, safety, reliability, and manufacturability of the U.S. nuclear stockpile.

The RS/6000 Enterprise Server Models S70, S70 Advanced, and S80 can also function as SP nodes, using attachments provided by the SP Switch or the LAN. These servers' exceptional performance is especially impressive for online transaction processing applications. With their large, single-node data capacity, they are also well-suited to the tasks associated with enterprise resource planning. Plus, their excellent query capability is ideal for today's business intelligence applications. These characteristics combined make the S70, S70 Advanced, and S80 servers an excellent third-tier choice for data storing and serving in three-tier e-business environments where POWER3 SMP nodes or 332 MHz SMP nodes are used as the middle tier.

Effective parallel computing requires high-bandwidth, low-latency internode communications. The SP Switch, a state-of-the-art IBM innovation, provides a maximum bidirectional data-transfer rate of over 120 MBps between each node pair.

The SP Switch Router is a high-performance I/O gateway that provides the fastest available means of communication between the SP system and the outside world or among multiple SP systems. This SP gateway combines the Ascend GRF with the IBM SP Switch Router adapter to enable direct network attachment to the SP Switch. Other media cards connect to a variety of standard external networks. Each media card has its own hardware engine, enabling SP I/O to scale nearly one-to-one with the number of cards.

The SP system can also scale disk I/O nearly one-to-one with processors and memory making access to terabytes of data possible and expansions or upgrades easier to manage. If you outgrow your existing system, you can readily add increments of computing power.

The SP system delivers balanced performance with processor, memory, switch, and I/O scalability. Over time, the SP has demonstrated leadership in standard industry benchmarks. An SP e-business server is recognized by the 1998 Guinness Book of Records for an Internet volume of 110,414 hits in one minute recorded at the Nagano Winter Olympic Games.

1.1.1 Origins

In the late 1980s, IBM set out to build a supercomputer for large, technical customers. The High Performance Supercomputer Systems Development Laboratory (HPSSDL) was formed within the IBM Large Systems Division in Kingston and Poughkeepsie, New York. HPSSDL intended to create a supercomputer with a more familiar personality, one based on widespread, non-exotic technology. Not surprisingly, the IBM ES/9000 mainframe vector processor architecture initially provided the basis for development. This

architecture eventually proved to be too limiting. Implementation aspects, such as guaranteed instruction execution order, special interrupt handling, and a unique floating point representation (incompatible with the emerging IEEE-based standard) restricted the speed and interoperability of the design.

In 1990, the IBM advanced workstation division in Austin, Texas introduced the RISCSystem/6000 (RS/6000) family of UNIX-based workstations and servers. These early RS/6000 machines boasted stellar floating point performance for their time, owing to the strength of the Performance Optimization with Enhanced RISC (POWER) CPU architecture. The fact that they ran UNIX was of great interest to HPSSDL, as UNIX was entrenched in their target market of large scientific and technical customers. HPSSDL, which was at an impasse with mainframe processor technology, experimented with off-the-shelf RS/6000 machines by adding ESCON adapters and interconnecting them with an ESCON Director. The RS/6000 machines were repackaged as nodes and placed in frames. Only five of the large, sheet metal drawers for the nodes could be placed in one frame. The drawers were judged to be too small to contain a person, so they were nicknamed *dog coffins*.

As HPSSDL was creating dog coffins, an IBM research group in Yorktown, New York was working on a high-speed switch code-named Vulcan. Yet another group in Yorktown was trying to solve the problem of deploying these hot new RS/6000 workstations to the desktops of IBM workers, as well as dealing with the system management headaches that come with LAN administration. This group developed a frame that could house 16 RS/6000 machines, as well as management software called Agora, to create a true LAN-in-a-can.

In December 1991, these independent efforts began to come together. HPSSDL was reorganized and renamed HPSSL (the Development part of the name was dropped) under the leadership of Irving Wladawsky-Berger. (Mr. Wladawsky-Berger went on to become head of the RS/6000 Division and currently is the Vice President of Technology and Strategy for the Enterprise Systems Group.) HPSSL's mission was to ship a product in 12 months. Designing a system from scratch was out of the question given the time constraint; so, a task force was created to select the necessary system components from available IBM technology. The RS/6000 Model 370 furnished the node technology. The Yorktown LAN consolidators provided their knowledge and experience in packaging the nodes. The Vulcan switch, now code-named Envoy (the origin of the E commands for the switch), was chosen over the ESCON interconnect, which was experiencing problems with excessive latency. Work from the ESCON interconnect experiment formed the

basis for the first iteration of the Vulcan switch software. The total product was introduced to the marketplace as the SP1.

In September 1993, Argonne National Laboratories in Argonne, Illinois, received shipment of the first SP1, a 128-node system. Cornell University in Ithaca, New York, bought a 512-node system shortly thereafter. Next came the petroleum industry. By the end of the year, 72 systems had been installed around the world, all with scientific and technical customers.

Initially, IBM had no intention of positioning the SP1 in the commercial marketplace, but commercial customers started knocking on the door of IBM. In the early 1990s, the death of the mainframe was accepted as conventional wisdom. Therefore, many large commercial enterprises were looking for alternatives to the mainframe in order to deploy new applications. IBM formed an application solutions group for the SP1, which, among other things, ported a parallel version of Oracle's database to the SP1. In 1994, SP development absorbed personnel from the discontinued AIX/ESA product who bolstered the system's manageability and helped spawn the Parallel System Support Program (PSSP, described in Section 10.1, "Parallel System Support Programs (PSSP)" on page 181). By the end of 1994, the commercial segment accounted for 40 percent of all installed SPs. By the end of 1996, the share climbed to 70 percent.

The SP2 was announced in 1994. It incorporated new node types from Austin and a faster switch, code-named Trailblazer (the origin of the tb2 and tb3 nomenclature of the switch adapters). The SP2 had moved out from under the umbrella of the Large Systems Division and was its own little enterprise within IBM. SP2 sales were strong: 352 systems were installed by the end of 1994 and 1,023 by the end of 1995.

In 1996, the SP2 was renamed to simply the SP and formally became a product of the RS/6000 Division. It represents the high-end of the RS/6000 family. IBM secured a number of large SP contracts. One of particular note is the ASCI project of the US Department of Energy. These contracts, coupled with the broad marketplace acceptance of the product, have fueled SP development. In 1996, IBM introduced a faster version of the Trailblazer switch, more than doubling the bandwidth of its predecessor, new nodes, including Symmetric Multiprocessor (SMP) versions, and more robust and functional PSSP software.

1.1.2 Management and availability

Each SP node is managed by a fully functional AIX operating system, providing access to the thousands of available AIX applications.

In addition, the system can be partitioned into pools of nodes. For example, two nodes can work as a Lotus Notes server, while ten others process a parallel database.

The SP system optimizes high availability through built-in redundancy, subsystem recovery, component error checking and correction, RAID5, external and internal disk mirroring, and hardware and software monitoring. Clusters of up to 16 SP nodes are supported by one of the industry's leading software products for critical application backup and availability, High Availability Cluster Multi-Processing (HACMP) for AIX. If an error, such as a node failure, occurs, the system can execute a recovery script that transfers the work to another node and prevents the application from going down.

Managing large systems is always a complex process. For the SP system, including attached RS/6000 Enterprise Server Models S70, S70 Advanced, and S80, a single graphical operations console that displays hardware, software, job, and user status makes system management easier. The system administrator uses this console, an RS/6000 system known as the control workstation, and the Parallel Systems Support Programs (PSSP) software product (available with the SP system) to perform management tasks including user and password management and job accounting as well as system startup/shutdown, monitoring, and partitioning.

In addition, the SP system offers a wide range of open system management software tools for operations and administration, availability, deployment, and security management. Included are the Tivoli and NetView network management products, Tivoli Storage Manager for back up and recovery, and Performance Toolbox Parallel Edition for performance monitoring.

1.1.3 Features and benefits

The SP system is a general-purpose scalable parallel system based on share-nothing architecture. Generally available SP systems range from 2 to 128 processor nodes, and each processor node is functionally equivalent to a stand-alone RS/6000 workstation or server. Large SP systems with up to 512 nodes have been delivered and are successfully being used today. Each processor node contains its own copy of the standard AIX operating system and other standard RS/6000 system software. A set of new software products designed specifically for the SP allows the parallel capabilities of the SP to be effectively exploited.

The RS/6000 SP system takes the advantages of the industrial leading RS/6000 workstation technology, the latest RS/6000 processors are repackaged for use as SP nodes, and the SP nodes are interconnected by a

high-performance, multistage, packet-switched network for interprocessor communication in order to perform the parallel functions.

As a general-purpose parallel computer, SP systems are used productively in a wide range of application areas and environments in the high-end UNIX technical and commercial computing market. This broad-based success is attributable to the highly flexible and general-purpose nature of the system. RS/6000 SP system can be used for doing a variety of tasks including parallel, serial, batch, and interactive jobs.

The RS/6000 SP system is a member of the RS/6000 product family, and it is positioned in the top end of entire RS/6000 product line. The following are the features and benefits of the SP system:

POWER3, PowerPC, and POWER2 Super Chip

- Delivers the processing power required for large and complex applications
- Allows the flexibility to configure for optimum commercial or technical computing application performance

Open system design

- Supports many communication protocols, adapters, and peripherals for a flexible system

Configuration flexibility

- Provides various node types that can be intermixed on the system
- Supports various PCI and Micro Channel adapters

Multistage packet switch

- Supports high-performance communications between processor nodes.
- Maintains point-to-point communication time independently of node location.
- SP Switch Router provides fastest available communication between SP systems and external networks.

Scalability

- Makes upgrading and expansion easier and allows for transparent application growth

System partitioning

- Isolates application subsystems and enables concurrent use of production and test AIX systems

Single point of administrative control

- Makes system management easier with less expertise and time required for most tasks

Comprehensive system management

- Provides the tools required to install, operate, and control an SP system
- Maintains consistency with other enterprise AIX systems management tools

High availability

- Helps avoid costly downtime due to system outages and provides an optional backup control workstation

S70, S70 Advanced, and S80 attachment

- Provides SP users with excellent performance for online transaction processing, enterprise resource planning, business intelligence, and e-business applications.
- Attached server applications communicate as an SP node on the SP Switch or through a LAN; attached servers can be easily managed, along with SP nodes, from the SP control workstation.

AIX operating system

- Provides a wealth of multiuser communications and systems management technologies
- Complies with major industry standards
- Provides AIX binary compatibility, where most AIX 4 applications already running on other RS/6000 systems can run unmodified

1.1.4 Business solutions

Installed in thousands of customer locations worldwide, the SP system delivers solutions for some of the most complex, large technical and commercial computing problems.

Customer uses include: Mission-critical commercial computing solutions to address business intelligence applications, server consolidation, and collaborative computing comprised of Lotus Notes, Lotus Domino Server, Internet, intranet, extranet, and groupware application solutions.

The SP database and computation scalability, critical for business intelligence applications including data warehousing, has led to more than 40 installations of greater than a terabyte of data.

Recognized in the industry as a high-capacity and reliable Web server, the SP system is an ideal base for e-business applications. Over 100 companies and organizations worldwide use it to handle their Web sites.

Technical computing users, including corporations, universities, and research laboratories, use the SP system for leading-edge applications, such as seismic analysis, computational fluid dynamics, engineering analysis, and computational chemistry.

SP solutions can be categorized into the following areas:

Business intelligence

- Provides scalable database capacity with support of leading parallel databases including IBM DB2 UDB EEE, Oracle Enterprise Edition, and Informix Dynamic Server AD/XP
- Offers proven scalable performance with leadership TPC-D results
- Delivers mainframe interoperability for optimal data movement

e-business

- Scalable growth and single management console virtually eliminate server proliferation issues associated with the addition of new servers to support the increasing number of Internet services and the complex dynamic workloads characteristic of network computing.
- Flexible node partitioning options permit multiple logical computing tiers for web business; logic and database servers are supported in a single physical system while system investment is preserved.

Enterprise resource planning

- Provides LAN consolidation, allowing multiple systems in a two or three tier client/server environment to be managed as a single system
- Provides high availability computing using the IBM industry-leading HACMP to provide back up, recovery, and fault-tolerant computing for mission-critical applications
- Provides application consolidation among multiple nodes within a single SP system, allowing ERP and supply chain planning applications from multiple vendors to take on a single-system appearance

Server consolidation

- Helps reduce the complexities and costs of systems management, lowering total cost of ownership and allowing simplification of application service level management

- Leverages investment in hardware and software, allowing better sharing of resources and licenses and distributing idle cycles instead of hot spots
- Provides the infrastructure that supports improved availability, data sharing, and response time

Technical computing

- Supports batch, interactive, serial, and parallel processing
- Provides outstanding floating-point performance
- Leads the way by supporting industry initiatives, such as PVM, MPI, and HPF

1.2 Hardware components

This chapter provides an overview of the hardware components of the IBM RS/6000 SP system. The basic components of the RS/6000 SP system are:

- Frames
- Processor nodes (includes SP-attached servers)
- Extension nodes (includes SP Switch Routers)
- Switches
- A control workstation (a high availability option is also available)

These components connect to your existing computer network through a local area network (LAN) making the RS/6000 SP system accessible from any network-attached workstation.

Figure 1 on page 10 illustrates a sample configuration of these hardware components. This gives you a rough idea of how they are connected. Thin nodes and an SP Switch are mounted in a tall frame. The thin nodes, an SP-attached server, and an SP Switch Router are connected to the SP Switch. The thin nodes, the SP-attached server, and the SP Switch Router are connected to SP Ethernet interface of a control workstation. The thin nodes, the SP Switch, the tall frame, and the SP-attached server are connected to RS-232C interface of the control workstation.

More detailed information on configurations and connections is found in the rest of the book.

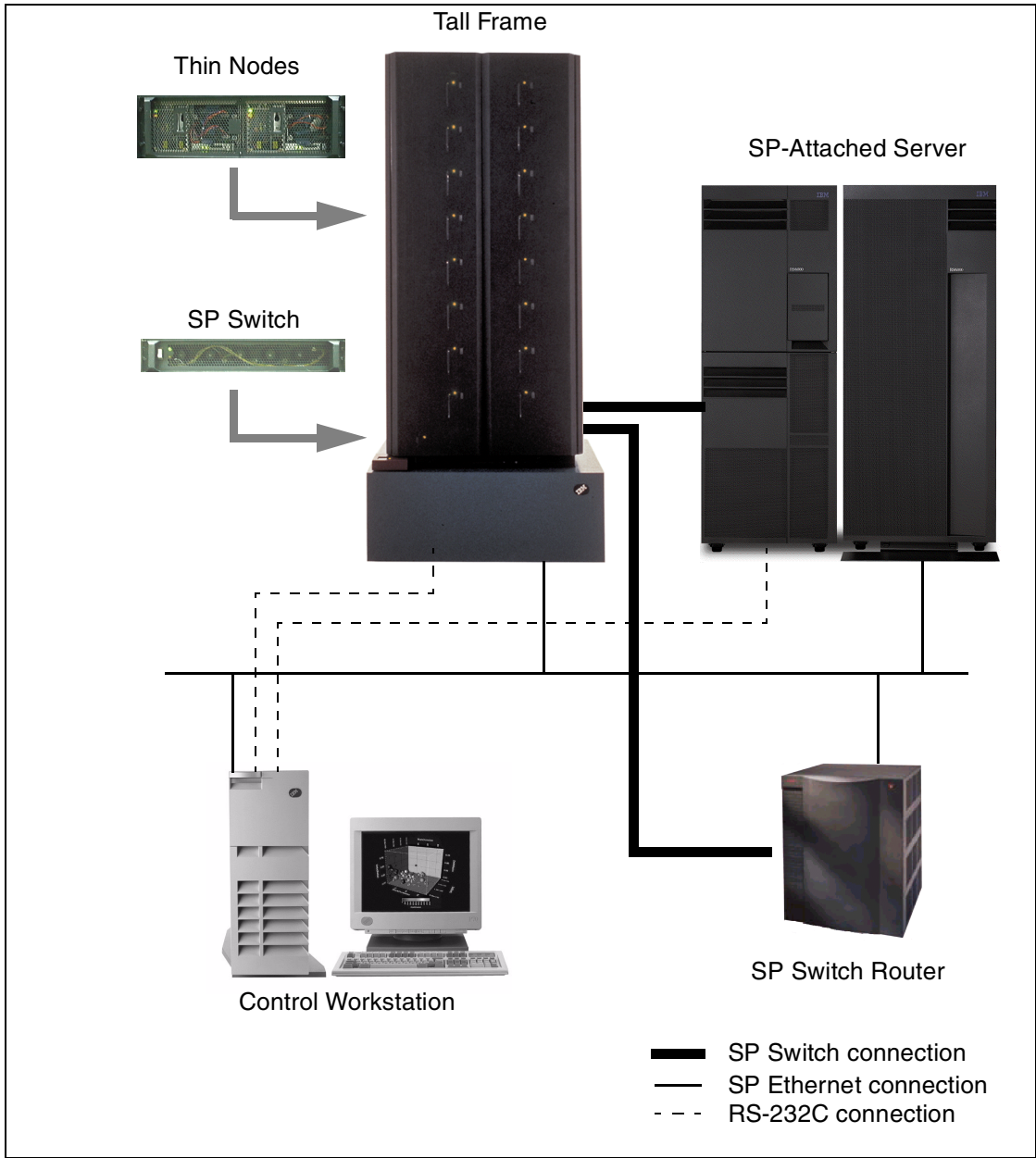


Figure 1. RS/6000 SP system sample configuration

1.2.1 Frames

IBM RS/6000 SP tall frames and all SEPBU power subsystems have been redesigned to accommodate the newest hardware features. These innovations were introduced through a simplified list of options for ordering SP frames. This listing has a total of five main frame options:

- Tall model frames
- Tall expansion frames
- Short model frames
- Short expansion frames
- SP Switch frames

The redesigned frames and power subsystems are completely compatible with all existing SP hardware.

Although the redesigned frames are completely compatible with all existing hardware, some hardware offerings are not compatible with the High Performance series of switches. These hardware items are:

- POWER3 SMP high nodes
- POWER3 SMP wide nodes
- POWER3 SMP thin nodes
- 332 MHz SMP wide nodes
- 332 MHz SMP thin nodes
- SP-attached servers
- SP Switch Routers
- SP Switches

If you are planning a system upgrade that includes any of these items, you must replace High Performance Switches with SP Switches.

For more information, refer to Chapter 2, "Frames" on page 15.

1.2.2 Processor nodes

The IBM RS/6000 SP System is scalable from 1 to 128 processor nodes, which can be contained in multiple SP frames. Up to 16 processor nodes can be mounted in a tall frame, while a short frame will hold up to eight processor nodes. SP systems consisting of more than 128 processor nodes are available. Consult your IBM representative for more information.

There are four types of RS/6000 SP processor nodes:

- Thin nodes
- Wide nodes
- High nodes
- SP-attached servers

1.2.2.1 Thin nodes

Thin nodes are available in a Symmetric MultiProcessor (SMP) configuration. All SMP thin nodes use Peripheral Component Interconnect (PCI) architecture.

The following nodes are currently available:

- 375 MHz POWER3 SMP nodes
- 332 MHz SMP nodes

1.2.2.2 Wide nodes

Wide nodes are available only in a SMP configuration. All SMP wide nodes use PCI architecture.

The following nodes are currently available:

- 375 MHz POWER3 SMP nodes
- 332 MHz SMP nodes

1.2.2.3 High nodes

High nodes occupy two full drawers, allowing four high nodes in a tall frame and two in a short frame. The high node is supported in both frame types with or without switch networks. Using a switch, up to 64 high nodes may be supported in tall frame SP systems.

The following nodes are currently available:

- 375 MHz POWER3 SMP nodes
- POWER3 SMP nodes

1.2.2.4 SP-attached servers

The SP-attached server is an IBM RS/6000 7017 Enterprise Server configured to operate with an RS/6000 SP System. These servers are available as either an S70 Enterprise Server, an S70 Advanced Server, or an S80 Enterprise Server. Each is a high-end, PCI-based, 64-bit, SMP unit that supports concurrent 32-bit and 64-bit applications.

Like a standard SP processor node, the SP-attached server can perform most SP processing and administration functions. However, unlike a standard SP processor node, the SP-attached server is housed in its own frame. Thus, the SP-attached server has both node-like and frame-like characteristics.

For more information, refer to Chapter 3, "Processor nodes" on page 29, or Chapter 4, "SP-attached servers" on page 73.

1.2.3 Extension nodes

An extension node is a non-standard node that extends the SP system's capabilities but that cannot be used in the same manner as a standard node.

Note that SP systems with extension nodes require PSSP 2.3 or later.

1.2.3.1 Dependent nodes

One type of extension node is a dependent node. A dependent node depends on SP nodes for certain functions but implements much of the switch-related protocol that standard nodes use on the SP Switch. Typically, dependent nodes consist of four major components as follows:

1. A physical dependent node - The hardware device requiring SP processor node support.
2. A dependent node adapter - A communication card mounted in the physical dependent node. This card provides a mechanical interface for the cable connecting the physical dependent node to the SP system.
3. A logical dependent node - It is made up of a valid, unused node slot and the corresponding unused SP Switch port. The physical dependent node logically occupies the empty node slot by using the corresponding SP Switch port. The switch port provides a mechanical interface for the cable connecting the SP system to the physical dependent node.
4. A cable - It connects the dependent node adapter with the logical dependent node. It connects the extension node to the SP system.

SP Switch Router

A specific type of dependent node is the IBM 9077 SP Switch Router. The 9077 is a licensed version of the Ascend GRF switched IP router that has been enhanced for direct connection to the SP Switch. These optional external devices can be used for high speed network connections or system scaling using HIPPI backbones or other communications subsystems, such as ATM or 10/100 Ethernet.

For more information, refer to Chapter 6, "SP Switch Routers" on page 97.

1.2.4 Switches

Switches provide a message-passing network that connects all processor nodes with a minimum of four paths between any pair of nodes. The SP series of switches can also be used to connect the SP system with optional external devices. A switch feature code provides you with a switch assembly and the cables to support node connections. The number of cables you receive depends on the type of switch you order.

There are two RS/6000 SP switch types as follows:

- SP Switches
- High Performance Switches

1.2.4.1 SP Switches

The SP Switch is available as either the 16-port SP Switch or the 8-port SP Switch-8:

- SP Switch2 (F/C 4012), 16-port switch
- SP Switch (F/C 4011), 16-port switch
- SP Switch-8 (F/C 4008), 8-port switch

1.2.4.2 High-performance switches

The High Performance Switches are being phased out and are only available for MES upgrades to existing systems. These switches are available as either the 16-port High Performance Switch or the 8-port HiPS LC-8 switch.

For more information, refer to Chapter 7, "SP Switch network" on page 109.

1.2.5 Control workstations

When planning your control workstation, you can view it as a server to the SP system applications. The subsystems running on the control workstation are the SP server applications for the SP nodes. The nodes are clients of the control workstation server applications. The control workstation server applications provide configuration data, security, hardware monitoring, diagnostics, a single point of control service, and optionally, job scheduling data and a time source.

For more information, refer to Chapter 8, "Control workstations" on page 123.

Chapter 2. Frames

The processor nodes can be mounted in either a tall or short SP frame. The frame spaces that nodes fit into are called drawers. A tall frame has eight drawers, while a short frame has four drawers. Each drawer is further divided into two slots. One slot can hold one thin node. A wide node occupies one drawer (two slots), and a high node occupies two drawers (four slots). An internal power supply is included with each frame. Frames get equipped with the optional processor nodes and switches that you order. Strictly speaking, there are three types of frames:

- Short frames
- Tall frames
- Switch frames

The tall and short frames are used to host nodes, and they are usually just called frames. The switch frames are used to host switches or Intermediate Switch Boards (ISB). This special type of frame can host up to eight switch boards. After the first SP was made commercially available some years ago, there have been a number of model and frame configurations. Each configuration was based on the frame type and the kind of node installed in the first slot. This led to an increasing number of possible prepackaged configurations when more nodes became available.

With the announcement on April 21, 1998, the product structure of RS/6000 SP system has been simplified. A frame was introduced. This frame replaces the old tall frame. The SP system was simplified with only two models (Model 500 and Model 550) and two expansion frame features (F/C 1500 and F/C 1550) compared with the six models and six features for each node type that were previously introduced.

Figure 2 shows the tall model frame (Model 550).



Figure 2. SP frame (model 550)

The most noticeable difference between the new and old tall frame is the reduction in height. Another physical difference is the footprint. Before this new frame offering, the frame and the first node in the frame were tied together forming a model. Each new node becoming available was potentially installable in the first slot of a frame; so, a new model was born. With the new offering, IBM simplified the SP frame options by decoupling the imbedded node from the frame offering. Therefore, when you order a frame, all you receive is a frame with the power supply units and a power cord. All nodes, switches, and other auxiliary equipment are ordered separately.

All new designs are completely compatible with all valid SP configurations using older equipment. Also, all new nodes can be installed in any existing SP frame provided that the required power supply upgrades have been implemented in that frame. The reason for this is that the SMP nodes have higher power consumption. Therefore, there is a higher power requirement for the frame.

The Model 500 is a short frame and the Model 550 is a tall frame. These models can house a mix of thin, wide, or high nodes. The Model 550 has a height of 75.8 inches (1.9 m), which provides an easier fit for the entrance doors than the old 79 inches (2.01 m) tall frame. However, it requires more floor space because the depth of the tall frame is 51 inches (1.3 m), which is 7 inches (0.2 m) greater than the old tall frame. The Model 500 has the same dimensions as the old short frame.

2.1 Short frames

Short frames are now available in two variations:

1. Short model frames (Model 500)
2. Short expansion frames (F/C 1500)

All short frames (including previous feature codes) are 1.25 m in height. However, the short frame Scalable Electric Power Base Unit (SEPBU) power subsystem has been redesigned to accommodate the latest SP processor nodes. This upgrade is required before you can use SMP wide nodes and SMP thin nodes in older frames.

Even with the redesigned power subsystem, these frames are completely compatible with all existing SP systems. All existing node and SP switch types can be directly installed as long as configuration rules are not violated. However, the High Performance series of switches is being phased out, and these switches are not compatible with some SP systems.

As in the older frames, the redesigned short frame has four drawers and can, therefore, house up to eight thin nodes, four wide nodes, or two high nodes. Short frames also accommodate an eight port switch and all power subsystems. All node types can be directly installed as long as configuration rules are not violated. All frames are designed for concurrent maintenance, where each processor node can be removed and repaired without interrupting operations on other nodes.

As in the older frames, redundant power (F/C 1213) is an option with the redesigned short frame SEPBU. With this option, if one power supply fails, another takes over. These power supplies are self regulating SEPBU units that have been upgraded to meet the increased power demand of the new SMP nodes. SEPBU's with the N+1 feature are also designed for concurrent maintenance. If a power book fails, it can be removed and repaired without interrupting running processes on the nodes.

IBM simplified the SP frame options by decoupling the imbedded node from the frame offering. Therefore, when you order a frame, all you receive is a frame with SEPBU and power cord. All nodes, switches, and other auxiliary equipment are ordered separately.

Attention

In order to maintain your entire SP system at the same electrical potential, you must attach a frame-to-frame ground between all frames in your SP system using IBM cables (P/N 46G5695).

Figure 3 illustrates a short frame from a front view.

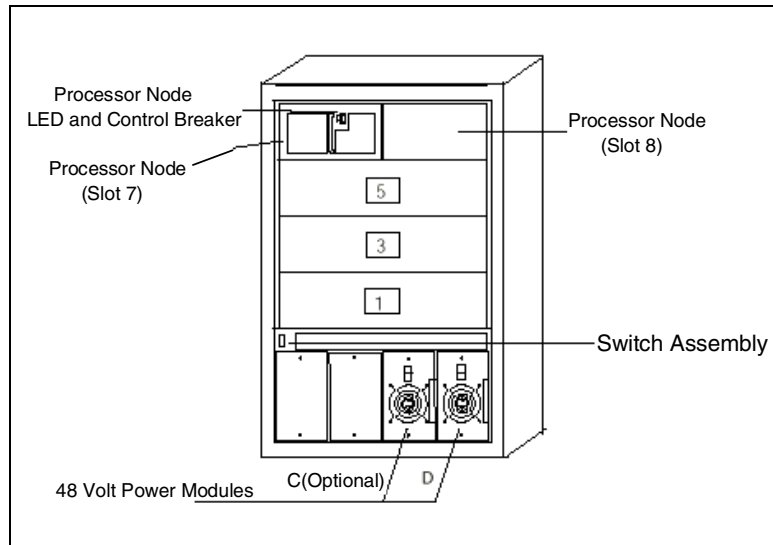


Figure 3. Front view of short frame

2.1.1 Short model frames (Model 500)

The model frame is always the first frame in an SP system and it designates the type or model class of your SP system. The base level Model 500 SP system has a short frame with four empty node drawers and a 5.0 kW single phase SEPBU power subsystem.

All processor nodes and the optional switch must be purchased separately for these 1.25 m (49 inch) frames. One SP Switch-8, and up to eight thin nodes, four wide nodes or two high nodes, may be installed in the Model 500 frame. Other frames that you connect to the model frame are known as expansion frames.

2.1.1.1 Non-switched configuration

This configuration consists of 1 to 8 processor nodes mounted in one required Model 500 frame and up to three additional non-switched (short) expansion frames (F/C 1500).

2.1.1.2 Switched configuration

This configuration consists of one to eight processor nodes connected through a single SP Switch-8 (8-port switch, F/C 4008). These nodes are mounted in one required Model 500 frame containing the SP Switch-8 and in up to three additional non-switched (short) expansion frames (F/C 1500). In this configuration, each node requires either an SP Switch Adapter (F/C 4020), an SP Switch MX2 Adapter (F/C 4023), or the withdrawn SP Switch MX Adapter (F/C 4022).

Nodes in the non-switched expansion frames (F/C 1500) share unused switch ports in the model frame. When short frames are used in a switched configuration, only the Model 500 frame can be equipped with a switch. SP switches cannot be mounted in the F/C 1500 frames.

2.1.2 Short expansion frames (F/C 1500)

F/C 1500 is the base offering for the 1.25 m (49 inch) SP expansion frame. These frames are equipped with a 5.0 kW single phase SEPBU self-regulating power subsystem. All 1.25 m frames have four empty node drawers for separately purchased nodes. Up to eight thin nodes, four wide nodes, or two high nodes may be installed in these frames. Switches cannot be mounted in F/C 1500 expansion frames.

You must populate each expansion frame with optional nodes as permitted by system configuration rules. These configuration rules impose limits on the number and location of each type of node that may be included in each system. The configuration rules will also vary depending on how your SP model frame was configured.

F/C 1500 expansion frames can only be configured with processor nodes. Expansion frames that are configured with processor nodes only (no switches) are known as non-switched expansion frames.

2.1.2.1 Using non-switched expansion frames

Model 500 SP systems can be fully utilized using F/C 1500 expansion frames. Model 500 systems have a capacity for up to eight nodes. If you fill the model frame before installing all eight nodes, you can install additional nodes in the system by using F/C 1500 non-switched expansion frames. The model frame may be either:

- Configured with processor nodes and a switch
- Configured with processor nodes only

Configurations with processor nodes and a switch

Non-switched expansion frames may be added to the Model 500 SP frame configured with processor nodes and an SPS-8 Switch to take advantage of unused switch ports.

One example of an underutilized switch would be a Model 500 frame with four wide nodes and an SP-8 Switch. In this case, the frame is fully populated yet only four of the eight switch ports are used. In this case, you can add non-switched expansion frames to the model frame to take advantage of the four unused switch ports. In these systems, node-to-node data transfers are completed through the switch.

Attention

Nodes must be in sequence in this system configuration. Empty node drawers are not allowed.

Configurations with processor nodes only

Non-switched expansion frames may be added to the Model 500 SP frame configured with processor nodes only (no switch) to take advantage of unused node slots. Model 500 systems have a capacity for up to eight nodes. If you fill the model frame by placing two high nodes in that frame, you can install six additional nodes in the system by using F/C 1500 non-switched expansion frames. In these systems, node-to-node data transfers are completed over the SP Ethernet.

2.2 Tall frames

Tall frames are now available in three variations:

1. Tall model frames (Model 550)
2. Tall expansion frames (F/C 1550)
3. SP Switch frames (F/C 2031)

Previous tall frames were 2.01 m in height; however, these frames have been redesigned so that they are now 1.93 m high. These frames are completely compatible with all existing SP systems. All existing node and SP Switch types can be directly installed as long as configuration rules are not violated.

However, the High Performance series of switches is being phased out, and these switches are not compatible with some SP systems.

As in the older frame designations, the redesigned tall frames have eight drawers and can house up to 16 thin nodes, eight wide nodes, or four high nodes; all node types can be mixed in these frames. These frames will also accommodate a switch and all power subsystems. All frames are designed for concurrent maintenance, where each processor node can be removed and repaired without interrupting operations on other nodes.

In addition to the reduced height, tall frames now contain an upgraded SEPBU power subsystem to accommodate the latest SP processor nodes. This upgrade is required before you can use SMP wide nodes and SMP thin nodes in older frames.

As in the older frames, the upgraded SEPBU power subsystem comes equipped with redundant (N+1) power supplies. If one power supply fails, another takes over. These power supplies are self-regulating SEPBU units designed for concurrent maintenance, where a failed power book can be removed and repaired without interrupting running processes on the nodes.

IBM simplified the SP frame options by decoupling the imbedded node from the frame offering. Therefore, when you order a frame, all you receive is a frame with SEPBU and power cord. All nodes, switches, and other auxiliary equipment are ordered separately.

Attention

In order to maintain your entire SP system at the same electrical potential, you must attach a frame-to-frame ground between all frames in your SP system using IBM cables (P/N 46G5695).

Figure 4 illustrates tall frame from front and rear views.

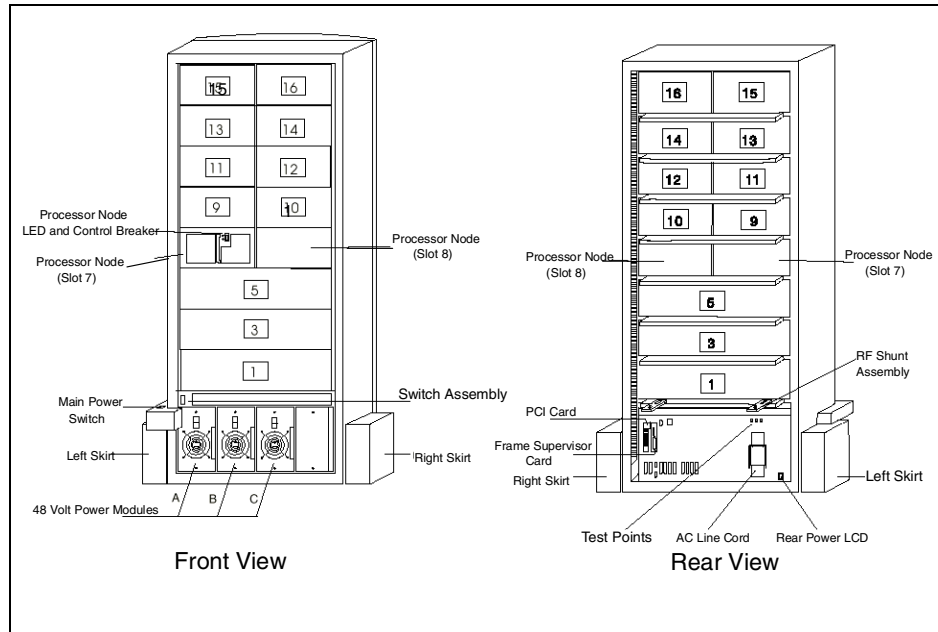


Figure 4. Front and rear views of tall frame

2.2.1 Tall model frames (Model 550)

The model frame is always the first frame in an SP system and it designates the type or model class of your SP system. The base level Model 550 SP system has a tall frame with eight empty node drawers and a 10.5 kW three phase SEPBU power subsystem.

All processor nodes and optional switches must be purchased separately for these 1.93 m (75.75 inch) frames. Either one SP Switch or one SP Switch-8 and up to sixteen thin nodes, eight wide nodes, or four high nodes, may be installed in these frames. Other frames that you connect to the model frame are known as expansion frames.

2.2.1.1 Non-switched configuration

This configuration consists of 1 to 64 processor nodes mounted in one required Model 550 frame and in additional non-switched (tall) expansion frames (F/C 1550).

2.2.1.2 SP Switch-8 configuration

This configuration consists of 1 to 8 processor nodes mounted in one required Model 550 frame equipped with an SP Switch-8 (8-port switch, F/C 4008). Non-switched expansion frames (F/C 1550) are supported in this configuration only if the model frame is filled before the total node count of eight is reached. In this configuration, each node requires either an SP Switch Adapter (F/C 4020), an SP Switch MX2 Adapter (F/C 4023), or the withdrawn SP Switch MX Adapter (F/C 4022). Nodes in the non-switched expansion frames share unused switch ports in the model frame.

2.2.1.3 Single-stage SP Switch configuration

This configuration consists of 1 to 80 processor nodes mounted in one required Model 550 frame equipped with an SP Switch (16-port switch, F/C 4011). Depending on the number of nodes in your system, up to five switched expansion frames (five F/C 1550 plus five F/C 4011) may be added to the system.

Single-staged system configurations can also utilize non-switched expansion frames (F/C 1550). Nodes in the non-switched expansion frames share unused switch ports that may exist in the model frame and in the switched expansion frames.

In single-stage switch configurations, all nodes require either an SP Switch Adapter (F/C 4020), an SP Switch MX2 Adapter (F/C 4023), or the withdrawn SP Switch MX Adapter (F/C 4022). No more than 64 of the 80 nodes in these systems may be high nodes.

2.2.1.4 Two-stage SP Switch configuration

The standard two stage switch configuration has 65 to 128 processor nodes. No more than 64 of the 128 nodes may be high nodes. All nodes in the system must have either an SP Switch Adapter (F/C 4020), an SP Switch MX2 Adapter (F/C 4023), or the withdrawn SP Switch MX Adapter (F/C 4022).

These nodes are mounted in one required Model 550 frame equipped with an SP Switch (16-port switch, F/C 4011) and in switched expansion frames (F/C 1550 plus F/C 4011). The SP Switches in these frames form the first switching layer.

This system configuration also requires an SP Switch frame (F/C 2031,) which forms the second switch layer. The second stage switches in the SP Switch frame are used for high performance parallel communication between the SP Switches mounted on the model and switched expansion frames.

Switch traffic is carried through concurrent data transmissions using the Internet protocol (IP).

Two-stage switch system configurations can also utilize non-switched expansion frames (F/C 1550). Nodes in the non-switched expansion frames share unused switch ports that may exist in the model frame and in the switched expansion frames.

Attention

Alternate two-stage switch configurations mounting fewer than 65 nodes or more than 128 nodes are available. Two-stage configurations using less than 65 nodes are simpler to scale up than single stage switch configurations.

2.2.2 Tall expansion frames (F/C 1550)

F/C 1550 is the base offering for the 1.93 m (75.8 inch) SP expansion frame. These frames are equipped with a 10.5 kW three phase SEPBU self-regulating power subsystem. All 1.93 m frames have eight empty node drawers for separately purchased nodes. Up to sixteen thin nodes, eight wide nodes, or four high nodes may be installed in these frames.

You must populate each expansion frame with optional SP Switches and nodes as permitted by system configuration rules. These configuration rules impose limits on the number and location of each type of node and switch that may be included in each system. The configuration rules will also vary depending on how your SP model frame was configured.

There are two standard configurations for F/C 1550 expansion frames. They are:

1. An expansion frame configured with processor nodes only.
 - This configuration is known as a non-switched expansion frame.
2. An expansion frame configured with processor nodes and an SP Switch.
 - This configuration is known as a switched expansion frame.

2.2.2.1 Using non-switched expansion frames

A non-switched expansion frame is defined as a base offering expansion frame equipped with processor nodes only. Some Model 550 SP system configurations can be scaled into larger systems using these frames. These SP system configurations are:

- Configurations using Model 550 frames equipped with processor nodes and a switch.
- Configurations using switch configured Model 550 frames and F/C 1550 expansion frames equipped with processor nodes and a switch (switched expansion frames).
- Model 550 frames equipped with processor nodes only.

Configurations with processor nodes and a switch

Non-switched expansion frames are added to SP frames configured with processor nodes and a switch to take advantage of unused switch ports resulting from certain system configurations. These unused switch ports may be in the model frame or in a switched expansion frame. In these systems, the switch, which may have ports to attach up to 16 nodes, is not fully utilized.

One example of an under-utilized switch would be a tall frame with eight wide nodes and an SP Switch. In this example, the frame is fully populated, yet only eight of the sixteen switch ports are used. In this case, you can add non-switched expansion frames to the switch configured frame to take advantage of the eight unused switch ports. In these systems, node-to-node data transfers are completed through the switch.

Attention

If the switch used in this configuration is an SP Switch-8, the nodes must be placed sequentially in this system configuration. Empty node drawers are not allowed. If the switch is an SP Switch (16-port), nodes may be placed in any order and empty drawers are allowed.

Configurations with processor nodes only

Non-switched expansion frames may be added to the Model 550 SP frame configured with processor nodes only (no switch) to take advantage of unused node slots. In these systems, node-to-node data transfers are completed over the SP Ethernet.

2.2.2.2 Using switched expansion frames

A switched expansion frame is defined as a base offering expansion frame equipped with processor nodes and a switch. These frames are added to SP systems with switch configured Model 550 frames. Configuration rules permit you to attach up to five switched expansion frames to these model frames. In some system configurations, you may have unused switch ports on either the model frame or on the switched expansion frames. Those unused switch

ports can be used with non-switched expansion frames to complete your system.

If your SP system uses single stage switching, you can scale your SP system into a system containing up to 80 nodes. See 2.2.1.3, "Single-stage SP Switch configuration" on page 23 for more information.

If your SP system uses two stage switching, your SP systems can be scaled even larger. In these systems, 128 nodes (or more) are supported. See 2.2.1.4, "Two-stage SP Switch configuration" on page 23 for more information.

2.3 Switch frames

An SP Switch frame is defined as a base offering tall frames equipped with either four or eight SP Switches. These frames do not contain processor nodes. SP Switch frames are used to connect model frames and switched expansion frames that have maximized the capacity of their integral switch. Switch frames can only transfer data within the local SP system.

The base level SP Switch frame (F/C 2031) contains four SP Switches. An SP Switch frame with four SP Switches will support up to 128 nodes.

The base level SP Switch frame can also be configured into systems with fewer than 65 nodes. In this application, the SP Switch frame will greatly simplify future system growth. Consult your IBM sales representative for information about SP systems configured with more than 128 nodes and for information about using switch frames in smaller systems.

Attention

The SP Switch frame is required when the sixth SP Switch is added to the system, and it is a mandatory prerequisite for all large scale systems.

Figure 5 illustrates SP Switch frame from the front view.

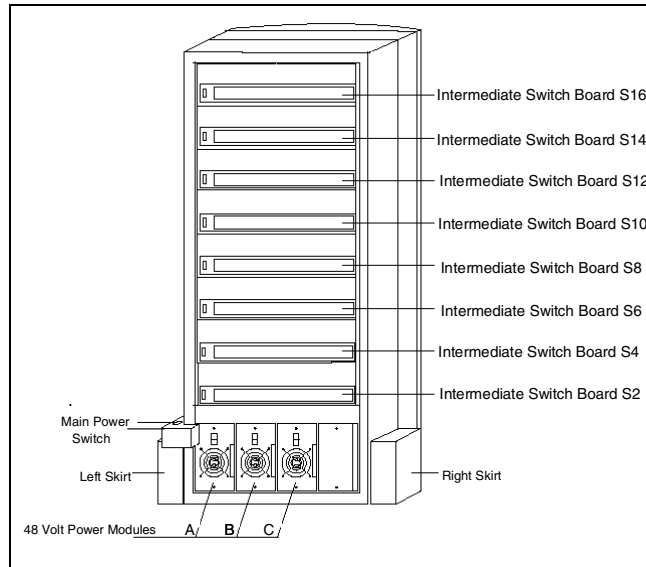


Figure 5. Front view of SP Switch frame with eight Intermediate switch boards

Chapter 3. Processor nodes

SP processor nodes are independent RS/6000 processors mounted in SP frames. Each node has its own processors, memory, disks, I/O slots, and AIX operating system. They are the basic building blocks of the SP systems. The only shared resource between the nodes is the interconnect networks, that is, either an SP Switch network or any other kinds of networks that are supported in the SP system. You can assign each SP nodes to work on a separate task or group the nodes for parallel applications.

There are three basic types of nodes: thin nodes, wide nodes, and high nodes. As the operating system and the major parts of the SP nodes are the same with a standard RS/6000 processor, thousands of existing applications can be run unchanged on SP nodes, which protects the customer investment to a large extent.

SP nodes are available in symmetric multiprocessing (SMP) configuration. All SMP nodes have Peripheral Component Interconnect (PCI) architecture. SMP nodes feature POWER3, POWER3-II, or PowerPC 604e microprocessors. These SP nodes are categorized as internal processor nodes. This chapter discusses these nodes. Figure 6 shows the 332 MHz thin nodes.

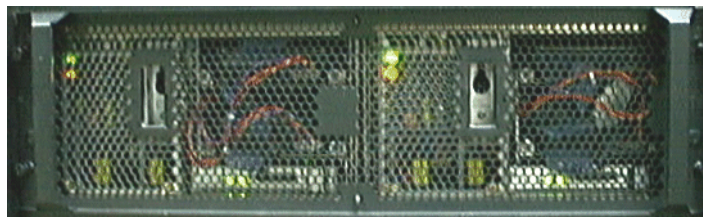


Figure 6. SP nodes (332 MHz thin node)

With the announcement of PSSP 3.1, the Enterprise Server S70, S70 Advanced, and S80 can also be integrated into the SP environment. These powerful RS/6000 SMP machines can act as standard SP nodes and provide excellent performance for online transaction processing, enterprise resource attachment planning, business intelligence, and e-business applications. Integrated Enterprise Servers are categorized as external processor nodes. Chapter 4, “SP-attached servers” on page 73, discusses these nodes.

3.1 POWER3-II SMP nodes

SP 375 MHz POWER3-II SMP node system design is based on the IBM PowerPC Architecture and the RS/6000 Platform Architecture. The node is designed as a bus-based symmetrical multiprocessor (SMP) system, using a 64-bit address and a 128-bit data system bus running at a 4:1 processor/clock ratio. Attached to the system bus (6xx bus) are from 2 to 4 PowerPC 630+ microprocessors, and a two chip memory-I/O controller.

The memory-I/O controller is a general purpose chip set that controls memory and I/O for systems, such as the POWER3-II SMP node, which implement the PowerPC MP System Bus (6xx bus). This chip set consists of two semi-custom CMOS chips, one for address and control, and one for data flow. The memory-I/O controller chip set includes an independent, separately-clocked "mezzanine" bus (6xx-MX bus) to which three PCI bridge chips and the SP Switch MX2 Adapter are attached. The POWER3-II SMP system architecture partitions all the system logic into the high speed processor-memory portion and to the lower speed I/O portion. This design methodology removes electrical loading from the wide, high-speed processor-memory bus (6xx bus) allowing this bus to run much faster. The wide, high-speed 6xx bus reduces memory and intervention latency while the separate I/O bridge bus supports memory coherent I/O bridges on a narrower, more cost-effective bus.

Figure 7 shows the POWER3-II system architecture block diagram.

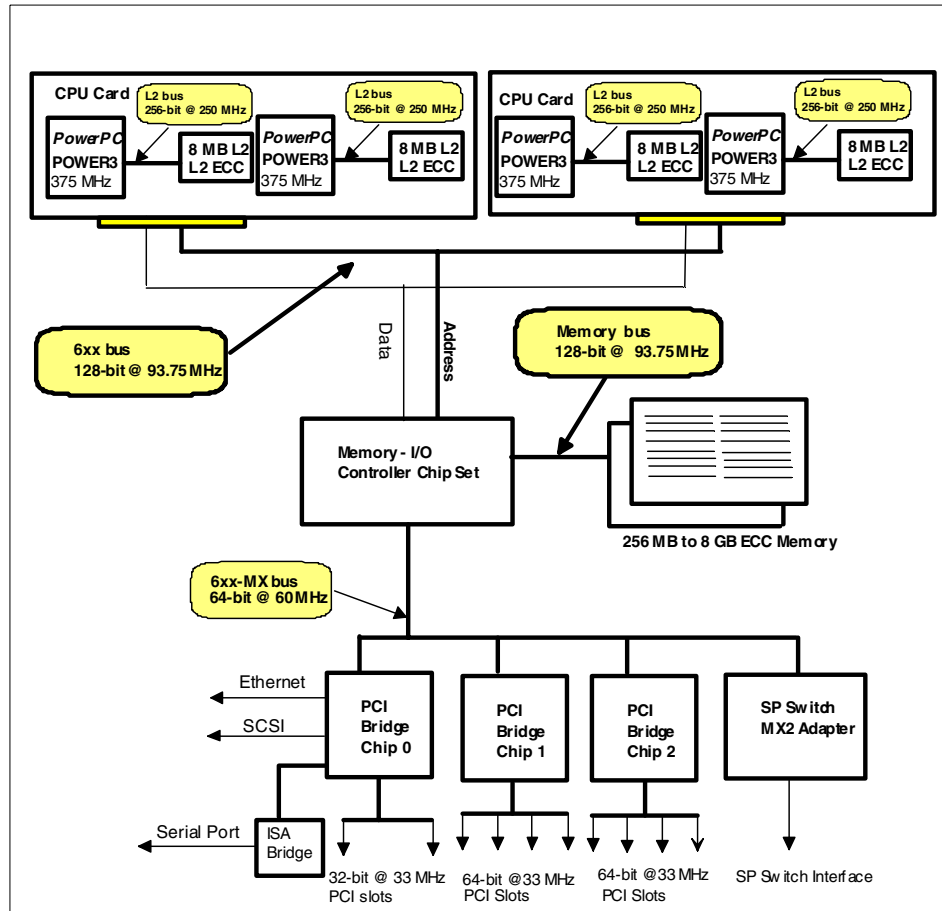


Figure 7. POWER3-II SMP node architecture

3.1.1 POWER3-II microprocessor

The POWER3-II design contains a superscalar core, which is comprised of eight execution units, supported by a high bandwidth memory interface capable of performing four floating-point operations per clock cycle. The POWER3-II design allows concurrent operation of fixed-point, load/store, branch, and floating-point instructions. There is a 32 KB instruction and 64 KB data level 1 cache integrated within a single chip in .22 um CMOS technology. Both instruction and data caches are parity protected.

The level 2 cache controller is integrated into the POWER3-II microprocessor with the data arrays and directory being implemented with external SRAM modules. The POWER3-II microprocessor has a dedicated external interface (separate from 6xx bus interface) for the level 2 cache accesses. Access to the 6xx bus and the level 2 cache can occur simultaneously. The level 2 cache is a unified cache (that is, it holds both instruction and data), and is configured for four-way set associative configuration. The external interface to the 8 MB of level 2 cache has 256-bit width and operates at 250 MHz. This interface is ECC protected.

The POWER3-II microprocessor is designed to provide high performance floating-point computation. There are two floating-point execution units, each supporting 3-cycle latency, 1-cycle throughput, and double/single precision Multiply-Add execution rate. Hence, the POWER3-II microprocessor is capable of executing four floating-point operations per clock cycle, which results in a peak throughput of 1500 MFLOPS.

3.1.1.1 System memory

The SP POWER3-II SMP system supports 256MB to 8 GB of 10ns SDRAM. System memory is controlled by the memory-I/O chip set through the memory bus. The memory bus consists of a 128-bit data bus and operates at 93.75 MHz clock cycle. As shown in Figure 1, this bus is separated from the System Bus (6xx bus), which allows for concurrent operations on these two buses. For example, cache-to-cache transfers can occur while a DMA operation is in progress to an I/O device. There are two memory card slots in the system. Each memory card contains 16 DIMM slots. 256 MB and 128 MB memory DIMMs are supported for GA. Memory DIMMs must be plugged in pairs and at least one memory card with minimum of 256 MB of memory must be plugged in for system to be operational. System memory is protected by Single Error Correction and Double Error Detection ECC code.

3.1.1.2 I/O subsystem

The memory-I/O controller chip set implements a 64-bit plus parity, multiplexed address and data bus (6xx-MX bus) for attaching three PCI bridge chips and the SP Switch MX2 Adapter. The 6xx-MX bus runs at 60 MHz concurrently and independently from the 6xx and memory buses. At 60 MHz clock cycle, the peak bandwidth of the 6xx-MX bus is 480 MBps. The three PCI bridge chips attached to 6xx-MX bus provides the interface for 10 PCI slots. Two 32-bit PCI slots are in thin node and eight additional 64-bit PCI slots are in wide node.

One of the PCI bridge chips (Bridge Chip0) provides support for integrated Ultra2 SCSI and 10Base2, 100BaseT Ethernet functions. The Ultra2 SCSI

interface supports up to four internal disks. An ISA bridge chip is also attached to PCI Bridge Chip0 for supporting two serial ports and other internally used functions in the POWER3-II SMP node.

3.1.1.3 System firmware and RTAS

The POWER3-II SMP node system firmware flash memory is located on the I/O planar. System firmware contains code that is executed by the POWER3-II microprocessor during the initial program load (IPL) phase of the system boot. It also supports various interactions between the AIX operating system and hardware. The extent and method of interaction is defined in the RS/6000 Platform Architecture (RPA). The Run Time Abstraction Software (RTAS) defined by RPA provides support for AIX and hardware for specific functions such as initialization, power management, time of day, I/O configuration, and capture and display of hardware indicators. RTAS and system IPL code are contained on 1 MB of flash memory.

3.1.1.4 System packaging

The POWER3-II SMP nodes use the same form factor as the original POWER3 SMP nodes.

Figure 8 shows the POWER-III SMP Thin node package.



Figure 8. POWER3-II SMP Thin node package

Figure 9 shows the POWER3-II SMP Wide node packaging.

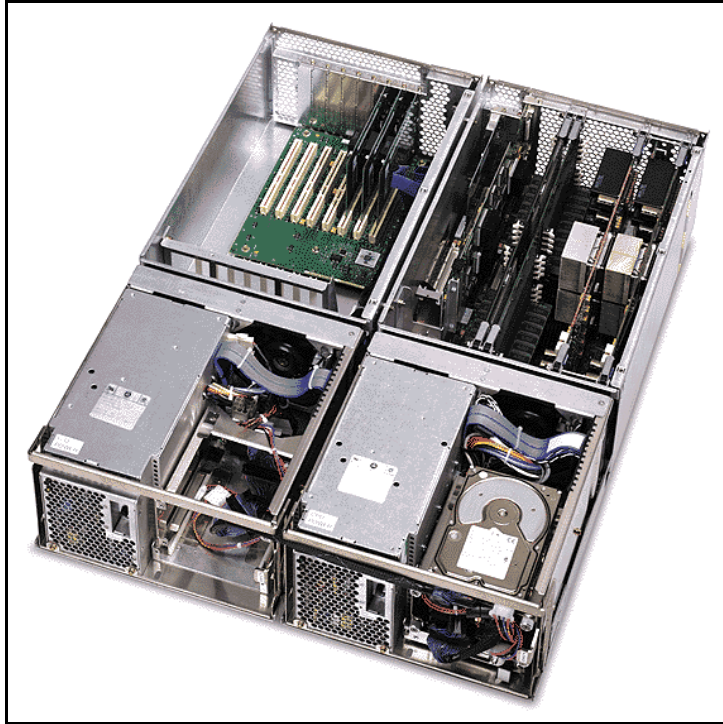


Figure 9. POWER3-II SMP Wide node package

3.2 POWER3 SMP High Node (F/C 2058)

375 MHz POWER3 SMP High Nodes (F/C 2058) use PCI bus architecture and have four, eight, twelve, or sixteen MHz 630FP 64-bit processors per node. Your IBM RS/6000 SP system must be operating at PSSP 3.2 (or later) to use these nodes.

The 375 MHz POWER3 SMP High Node occupies two full drawer locations, thus four nodes can be housed in a tall frame. These nodes require a 1.93 m tall, deep frame (Model 550) or expansion frame (F/C 1550); they are not supported in the withdrawn 2.01 m frame or in the 1.25 m frame. These nodes can be placed in the first node slot of a frame without requiring additional nodes.

Note

375 MHz POWER3 SMP High Nodes are not compatible with High Performance switches (F/C 4010 and F/C 4007)

3.2.1 PCI bus description

The 375 MHz POWER3 SMP High Node PCI bus contains one 32-bit and four 64-bit PCI slots for I/O adapters.

Additional PCI adapters can be attached to the bus by using up to six optional SP Expansion I/O Units. Each expansion unit has eight 64-bit PCI adapter slots.

3.2.2 Requirements

375 MHz POWER3 SMP High Nodes occupy two full node drawers. Up to four 375 MHz POWER3 SMP High Nodes may be installed in one tall/deep frame. The mandatory requirements are:

- PSSP 3.2 (or later) on the processor node, control workstation, and backup nodes
- Four processors (on one card, mounted in one slot)
- 1 GB of memory
- 9.1 GB mirrored DASD (with internal booting)

3.2.3 Options

Available options include the following:

- Four processors slots allowing a maximum of sixteen processors per node
- Four memory slots supporting up to 32 GB of memory
- Five PCI slots (four 64-bit and one 32-bit) for communication adapters
- A dedicated Mezzanine Bus (MX) slot for an optional switch adapter
- Integrated Ethernet with BNC and RJ45 ports (only one port can be used at a time):
 - 10Base2 Ethernet (BNC)
 - 10BaseT or 100BaseTX Ethernet (RJ45)
- Support for up to six SP Expansion I/O Units (F/C 2055)

- Two internal DASD bays supporting up to 72.8 GB of storage (36.4 GB mirrored)
- Integrated Ultra SCSI network
- External nine-pin RS-232 on the planar S2 port (supported only for HACMP serial heartbeat); a 9 to 25-pin converter cable is included with the node
 - Node-to-node HACMP cable (F/C 3124)
 - Frame-to-frame HACMP cable (F/C 3125)

3.2.4 Processor requirements and options

375 MHz POWER3 SMP High Nodes require a minimum of four 375 MHz, 630FP processors mounted on one card. You can order up to three additional four-processor cards (F/C 4350) to configure the node with a total of sixteen CPUs. Table 1 shows the processor options for the 375 MHz POWER3 SMP High Nodes.

Table 1. 375 MHz POWER3 SMP High Node (F/C 2058) processor options

F/C	Quantity	Description	Comments
4350	1 - 4	One processor card with four CPUs	1 required

3.2.4.1 Memory requirements and options

375 MHz POWER3 SMP High Nodes have one to four memory cards, a minimum of one GB (8 x 128 MB DIMMs) and a maximum of 32 GB (16 x 8 X 256 MB DIMMs). Groups of eight 128 MB and 256 MB DIMMs can be mixed on these memory cards. Table 2 shows the memory features for the 375 MHz POWER3 SMP High Nodes.

Table 2. 375 MHz POWER3 SMP High Node (F/C 2058) memory features

F/C	Description	Minimum Node Requirement	Maximum Allowed Per Node
4880	Base memory card	1	4
4133	(8) 128 MB DIMMs	1	16
4134	(8) 256 MB DIMMs	0	16

3.2.5 Disk requirements and options

375 MHz POWER3 SMP High Nodes can have one pair of internal DASD attached through an integrated Ultra SCSI network. The node can have either no internal DASD (with external booting) or from 9.1 GB to a maximum of 36.4 GB of mirrored, internal disk storage.

Optional internal disk devices are available as follows:

- 9.1 GB Ultra SCSI disk pair (F/C 2909)
- 18.2 GB Ultra SCSI disk pair (F/C 2918)
- 9.1 GB Ultra SCSI 10K RPM disk pair (F/C 3804)
- 18.2 GB Ultra SCSI 10K RPM disk pair (F/C 3810)
- 36.4 GB Ultra SCSI 10K RPM disk pair (F/C 3820)

External storage devices can be accessed through optional Ultra SCSI adapter (F/C 6207) and SSA adapter (F/C 6230).

3.2.6 Switch and communication adapter requirements and options

If you are planning to use a switch in your SP system, you need switch adapters to connect each RS/6000 SP node to the switch subsystem. SP switch adapters have several characteristics and restrictions, including the following.

Switch restrictions

375 MHz POWER3 SMP High Nodes are not supported with the SP Switch-8. You must use the SP Switch, 16-port (F/C 4011) or the SP Switch2, 16-port (F/C 4012).

Note

The 375 MHz POWER3 SMP High Node is not compatible with the older High Performance series of switches. If you install a POWER3 Wide Node into a switch-configured system, it must use only SP-type switches.

Switch adapters

The switch adapter for 375 MHz POWER3 SMP High Nodes does not occupy a PCI slot; it is installed into the Mezzanine (MX) bus. The MX bus connects the I/O planar with the system planar. Placing the switch adapter in the MX bus enables switch traffic to proceed at higher bandwidths and lower latencies.

For SP Switch systems, these nodes require the SP Switch MX2 Adapter (F/C 4023).

For SP Switch2 systems, these nodes require the SP Switch2 Adapter (F/C 4025)

I/O adapters

The 375 MHz POWER3 SMP High Node has five PCI (Peripheral Component Interconnect) adapter slots. A full line of PCI adapters is offered for these nodes including:

- SCSI-2
- Ethernet
- Token Ring
- FDDI
- ATM
- Async
- Wide Area Network (WAN)
- SSA RAID5
- S/390 ESCON
- Serial HIPPI

Note

A 10Base2 or 10BaseT/100BaseTX Ethernet adapter for the SP Ethernet is integrated into the POWER3 High Node and does not occupy a PCI slot.

3.2.7 375 MHz POWER3 SMP Wide Node (F/C 2057)

375 MHz POWER3 SMP Wide Nodes (F/C 2057) have PCI bus architecture and either two or four 375 MHz 64-bit processors per node. These nodes are functionally equivalent to an IBM RS/6000 7044-270 workstation. Your IBM RS/6000 SP system must be operating at PSSP 3.1.1 (or later) to use these nodes.

The node occupies one full drawer, thus eight nodes can be housed in a tall frame. These nodes can be placed in the first node slot of a frame without requiring additional nodes.

For electromagnetic compliance, these nodes are housed in an enclosure (F/C 9930).

If you plan to install a 375 MHz POWER3 SMP Wide Node into an early style 201 m or 1.25 m frame, a power system upgrade is required. Once the power system upgrade is done, these nodes are fully compatible with all existing SP system hardware excepting High Performance switches.

3.2.7.1 Bus description

The 375 MHz POWER3 SMP Wide Node PCI bus contains two 32-bit slots and eight 64-bit PCI slots divided into three logical groups. The first slot group (slot I2 and I3) is composed of the two 32-bit slots on the CPU side of the node. The second and third group each contain four 64-bit PCI slots (slots I1-I4 and slots I5-I8) on the I/O side of the node. The I1 slot on the CPU side is reserved for the optional SP Switch MX2 Adapter.

3.2.7.2 Requirements

375 MHz POWER3 SMP Wide Nodes occupy one full node drawer. They can be asymmetrically configured for memory, DASD, and adapters. Up to eight of these nodes can be installed in a tall frame and up to four in a short frame.

The mandatory requirements are:

- PSSP 3.1.1 (or later) on the control workstation, backup nodes, and processor node
- Two processors (mounted in one slot)
- 256 MB of memory
- 4.5 GB of mirrored DASD (with internal booting)
- An upgraded power system on early-style frames

3.2.7.3 Options

Available options include the following:

- Two processor slots allowing a maximum of four processors per node
- Two memory slots supporting up to 8 GB of memory
- Ten PCI slots (two 32-bit and eight 64-bit) for communication adapters
- A dedicated Mezzanine Bus (MX) slot for an optional switch adapter
- Integrated Ethernet with BNC and RJ45 ports (only one per port can be used at a time)
 - 10Base2 Ethernet (BNC)
 - 10BaseT Ethernet or 100BaseTX Ethernet (RJ45)

- Four DASD bays supporting up to 109.2 GB of Storage (54.6 GB mirrored)
- Integrated Ultra SCSI
- Standard service processor
- External nine-pin RS232 on the planar S2 port (supported only for HACMP serial heartbeat); a 9 to 25-pin converter cable is included with the node
 - Node-to-node HACMP cable (F/C 3124)
 - Frame-to-frame HACMP cable (F/C 3125)

3.2.7.4 Processor requirements and options

375 MHz POWER3 SMP Wide Nodes require a minimum of two processors mounted on one card. Optionally, you can order an additional processor card for a total of four CPUs. Table 3 shows the processor options for the 375 MHz POWER3 SMP Wide Nodes.

Table 3. 375 MHz POWER3 SMP Wide Node (F/C 2057) processor options

Feature Code	Quantity	Description	Comments
4444	1-2	One processor card with two CPUs	One required

3.2.7.5 Memory requirements and options

375 MHz POWER3 SMP Wide Nodes have two memory cards and require a minimum of 256 MB of memory. These nodes support a maximum of 8 GB of memory. Memory is supplied by 128 or 256 MB DIMMs mounted in pairs. The memory cards are not required to be configured symmetrically. Each card has a 4 GB capacity, with 8 GB addressable per node. Note that with the minimum memory installed (256 MB), the second card contains no DIMMs. Memory cards and DIMMs are not interchangeable between SMP and non-SMP nodes. Memory cards are not interchangeable between 332 MHz and 375 MHz POWER3 SMP nodes. Table 4 shows the memory features for the 375 MHz POWER3 SMP Wide Nodes.

Table 4. 375 MHz POWER3 SMP Wide Node (F/C 2057) memory features

Feature Code	Description	Minimum Node Requirement	Maximum Allowed Per Node
4098	Base Memory Card	2	2
4110	One Pair of 128 MB DIMMs (256 MB total)	1 pair	16 pairs

Feature Code	Description	Minimum Node Requirement	Maximum Allowed Per Node
4119	One Pair of 256 MB DIMMs (512 MB total)	n/a	16 pairs

3.2.7.6 DASD requirements and options

375 MHz POWER3 SMP Wide Nodes can have up to two pairs of internal DASD attached through an integrated Ultra SCSI network. This node can have either no internal DASD (with external booting) or from 4.5 GB to a maximum of 54.6 GB of mirrored internal disk storage.

Optional direct access storage devices are available as follows:

- 4.5 GB Ultra SCSI disk pair (F/C 2904)
- 9.1 GB Ultra SCSI disk pair (F/C 2909)
- 18.2 GB Ultra SCSI disk pair (F/C 2918)
- 9.1 GB Ultra SCSI 10K RPM disk pair (F/C 3804)
- 18.2 GB Ultra SCSI 10K RPM disk pair (F/C 3810)
- 36.4 GB Ultra SCSI 10K RPM disk pair (F/C 3820) - available only for I/O side DASD bays

Note

No special cables or adapters are required to mount these internal DASD. However, this node has an option (F/C 1241) that provides an independent SCSI hookup with the following characteristics:

- Eliminates the DASD controller as a single point of failure during mirroring
- Increases disk performance
- Balances disk loading

The F/C 1241 option requires a PCI type SCSI adapter F/C 6206

External storage devices can be accessed through optional Ultra SCSI adapter (F/C 6207), SCSI-2 adapter (F/C 6209), and SSA adapter (F/C 6230).

3.2.7.7 Switch and communication adapter requirements

If you are planning to use a switch in your SP system, you need switch adapters to connect each RS/6000 SP node to the switch subsystem. SP switch adapters have several characteristics and restrictions, including the following.

Switch adapters

The switch adapter for 375 MHz POWER3 SMP Wide Nodes does not occupy a PCI slot. The switch adapter for these nodes is installed into the Mezzanine (MX) bus. The MX bus connects the I/O planar with the system planar. Placing the switch adapter in the MX bus enables switch traffic to proceed at higher bandwidths and lower latencies.

In switch-configured systems, 375 MHz POWER3 SMP Wide Nodes require SP Switch MX2 adapter (F/C 4023).

Switch restrictions

375 MHz POWER3 SMP Wide Nodes are not compatible with the older High Performance series of switches. If you install this node into a switch-configured system, you must use an SP Switch or an SP Switch-8.

Switch adapters for these nodes are not interchangeable with either the switch adapters used on uniprocessor wide nodes or with the SP Switch MX adapter used on 332 MHz SMP nodes.

I/O adapters

The 375 MHz POWER3 SMP Wide Nodes has 10 PCI (Peripheral Component Interconnect) adapter slots. A full line of PCI adapters is offered for these nodes including:

- SCSI-2
- Ethernet
- Token Ring
- FDDI
- ATM
- Async
- Wide Area Network (WAN)
- SSA RAID5
- S/390 ESCON
- Serial HIPPI

Note

A 100BaseTX, 10BaseT, or 10Base2 adapter for the SP Ethernet is integrated into the node and does not occupy a PCI slot.

3.2.8 375 MHz POWER3 SMP Thin Node (F/C 2056)

375 MHz POWER3 SMP Thin Nodes (F/C 2056) have PCI bus architecture and either two or four 375 MHz 64-bit processors per node. These nodes are functionally equivalent to an IBM RS/6000 7044-270 workstation. Your IBM RS/6000 SP system must be operating at PSSP 3.1.1 (or later) to use these nodes.

The node occupies half of a drawer (one slot). Up to 16 of these nodes can be housed in a tall frame. When installed singly within a drawer, these nodes must be placed in an odd-numbered node slot. For complete information on node/frame configurations, see RS/6000 SP: Planning Volume 2, Control Workstation and Software Environment.

For electromagnetic compliance, these nodes are housed in an enclosure (F/C 9930). If you order a single node, a cover plate (F/C 9931) is included to fill the even-numbered slot opening.

If you plan to install a 375 MHz POWER3 SMP Thin Node into an early style 2.01 m or 1.25 m frame, a power system upgrade is required. Once the power system upgrade is done, these nodes are fully compatible with all existing SP system hardware except High Performance switches.

3.2.8.1 Bus description

The 375 MHz POWER3 SMP Thin Node PCI bus contains two 32-bit slots, I2 and I3). The I1 slot is reserved for the optional SP Switch MX2 Adapter.

3.2.8.2 Requirements

375 MHz POWER3 SMP Thin Nodes occupy one half node drawer. They can be asymmetrically configured for memory, DASD, and adapters. Up to sixteen of these nodes can be installed in a tall frame and up to eight in a short frame. The mandatory requirements are:

- PSSP 3.1.1 (or later) on the control workstation, backup nodes, and processor node.
- Two processors (mounted in one slot)
- 256 MB of memory

- 4.5 GB of mirrored DASD (with internal booting)
- An upgraded power system on early-style frames

3.2.8.3 Options

Available options include the following:

- Four processors in two slots
- Two memory slots supporting up to 8 GB of memory
- Two (32-bit) PCI slots for communication adapters
- A dedicated Mezzanine Bus (MX) slot for an optional switch adapter
- Integrated Ethernet with BNC and RJ45 ports (only one per port can be used at a time)
 - 10Base2 Ethernet (BNC)
 - 10BaseT Ethernet or 100BaseTX Ethernet (RJ45)
- Two DASD bays supporting up to 36.4 GB of Storage (18.2 GB mirrored)
- Integrated Ultra SCSI
- Standard service processor
- External nine-pin RS232 on the planar S2 port (supported only for HACMP serial heartbeat); a 9 to 25-pin converter cable is included with the node
 - Node-to-node HACMP cable (F/C 3124)
 - Frame-to-frame HACMP cable (F/C 3125)

3.2.8.4 Processor requirements and options

375 MHz POWER3 SMP Thin Nodes require a minimum of two processors mounted on one card. Optionally, you can order an additional processor card for a total of four CPUs. Table 5 shows the processor options for the 375 MHz POWER3 SMP Thin Nodes.

Table 5. 375 MHz POWER3 SMP Thin Node (F/C 2056) processor options

Feature Code	Quantity	Description	Comments
4444	1-2	One processor card with two CPUs	One required

3.2.8.5 Memory requirements and options

375 MHz POWER3 SMP Thin Nodes have two memory cards and require a minimum of 256 MB of memory. These nodes support a maximum of 8 GB of memory. Memory is supplied by 128 or 256 MB DIMMs mounted in pairs. The

memory cards are not required to be configured symmetrically. Each card has a 4 GB capacity, with 8 GB addressable per node. Note that with the minimum memory installed (256 MB), the second card contains no DIMMs. Memory cards and DIMMs are not interchangeable between SMP and non-SMP nodes. Memory cards are not interchangeable between 332 MHz and 375 MHz POWER3 SMP Nodes. Table 6 shows the memory features for the 375 MHz POWER3 SMP Thin Nodes.

Table 6. 375 MHz POWER3 SMP Thin Node (F/C 2056) memory features

Feature Code	Description	Minimum Node Requirement	Maximum Allowed Per Node
4098	Base Memory Card	2	2
4110	One Pair of 128 MB DIMMs (256 MB total)	1 pair	16 pairs
4119	One Pair of 256 MB DIMMs (512 MB total)	n/a	16 pairs

3.2.8.6 DASD requirements and options

375 MHz POWER3 SMP Thin Nodes can have one pair of internal DASD attached through an integrated Ultra SCSI network. The node can have either no internal DASD (with external booting) or from 4.5 GB to a maximum of 36.4 GB of mirrored internal disk storage.

Optional direct access storage devices are available as follows:

- 4.5 GB Ultra SCSI disk pair (F/C 2904)
- 9.1 GB Ultra SCSI disk pair (F/C 2909)
- 18.2 GB Ultra SCSI disk pair (F/C 2918)
- 9.1 GB Ultra SCSI 10K RPM disk pair (F/C 3804)
- 18.2 GB Ultra SCSI 10K RPM disk pair (F/C 3810)

Note

No special cables or adapters are required to mount these internal DASDs.

External storage devices can be accessed through optional Ultra SCSI adapter (F/C 6207), SCSI-2 adapter (F/C 6209), and SSA adapter (F/C 6230).

3.2.8.7 Switch and communication adapter requirements

If you are planning to use a switch in your SP system, you need switch adapters to connect each RS/6000 SP node to the switch subsystem. SP switch adapters have several characteristics and restrictions, including the following:

Switch adapters

The switch adapter for 375 MHz POWER3 SMP Thin Nodes does not occupy a PCI slot. The switch adapter for these nodes is installed into the Mezzanine (MX) bus. The MX bus connects the I/O planar with the system planar. Placing the switch adapter in the MX bus enables switch traffic to proceed at higher bandwidths and lower latencies.

In switch-configured systems, 375 MHz POWER3 SMP Thin Nodes require SP Switch MX2 adapter (F/C 4023).

Switch restrictions

375 MHz POWER3 SMP Thin Nodes are not compatible with the older High Performance series of switches. If you install this node into a switch-configured system, you must use an SP Switch or an SP Switch-8.

Switch adapters for these nodes are not interchangeable with either the switch adapters used on uniprocessor wide nodes or with the SP Switch MX adapter (F/C 4022) used on 332 MHz SMP nodes.

I/O adapters

The 375 MHz POWER3 SMP Thin Nodes has two PCI (Peripheral Component Interconnect) adapter slots. A full line of PCI adapters is offered for these nodes including:

- SCSI-2
- Ethernet
- Token Ring
- FDDI
- ATM
- Async
- Wide Area Network (WAN)
- SSA RAID5
- S/390 ESCON

Note

A 100BaseTX, 10BaseT, or 10Base2 adapter for the SP Ethernet is integrated into the node and does not occupy a PCI slot.

3.3 POWER3 SMP nodes

The POWER3 SMP node is the first scalable processor node that utilizes the POWER3 64-bit microprocessor. The floating-point performance of the POWER3 microprocessor makes this node an excellent platform for compute-intensive analysis applications. The POWER3 microprocessor offers technical leadership for floating-point applications by integrating two floating-point, three fixed-point, and two load/store units in a single 64-bit PowerPC implementation. Since the node conforms to the RS/6000 Platform Architecture, compatibility is maintained for existing device drivers, other subsystems, and applications. The POWER3 SMP node supports AIX operating systems beginning with version 4.3.2.

The POWER3 SMP node is available in two packages: Thin and wide. The thin node can accommodate up to two processor cards, two memory cards, and two PCI adapters. It also supports two Ultra SCSI hard files. The node is fully compliant with Revision 2.1 of the Peripheral Component Interconnect (PCI) specifications and implements three PCI buses. The first PCI bus, found in both thin and wide nodes, supports two 32-bit PCI slots running at 33 MHz. Integrated Ultra2 SCSI, Ethernet, and ISA bridge are also supported by the first PCI bridge chip. The other two PCI buses are located in the expansion I/O unit of the wide node. Each of these buses support four 64-bit PCI slots running at 33 MHz.

3.3.1 POWER3 SMP node system architecture

POWER3 SMP node system design is based on the IBM PowerPC Architecture and the RS/6000 Platform Architecture. The node is designed as a bus-based symmetrical multiprocessor (SMP) system using a 64-bit address and a 128-bit data system bus running at a 2:1 processor clock ratio. Attached to the system bus (6xx bus) are one to two PowerPC 630 microprocessors and a two chip memory-I/O controller.

The memory-I/O controller is a general purpose chip set that controls memory and I/O for systems, such as the POWER3 SMP node, which implement the PowerPC MP System Bus (6xx bus). This chip set consists of two semi-custom CMOS chips, one for address and control, and one for data

flow. The memory-I/O controller chip set includes an independent, separately-clocked mezzanine bus (6xx-MX bus) to which three PCI bridge chips and the SP Switch MX2 Adapter are attached. The POWER3 SMP system architecture partitions all the system logic into the high speed processor-memory portion and to the lower speed I/O portion. This design methodology removes electrical loading from the wide, high-speed processor-memory bus (6xx bus) allowing this bus to run much faster. The wide, high-speed 6xx bus reduces memory and intervention latency, while the separate I/O bridge bus supports memory coherent I/O bridges on a narrower, more cost-effective bus.

Figure 10 shows the POWER3 SMP node system architecture block diagram.

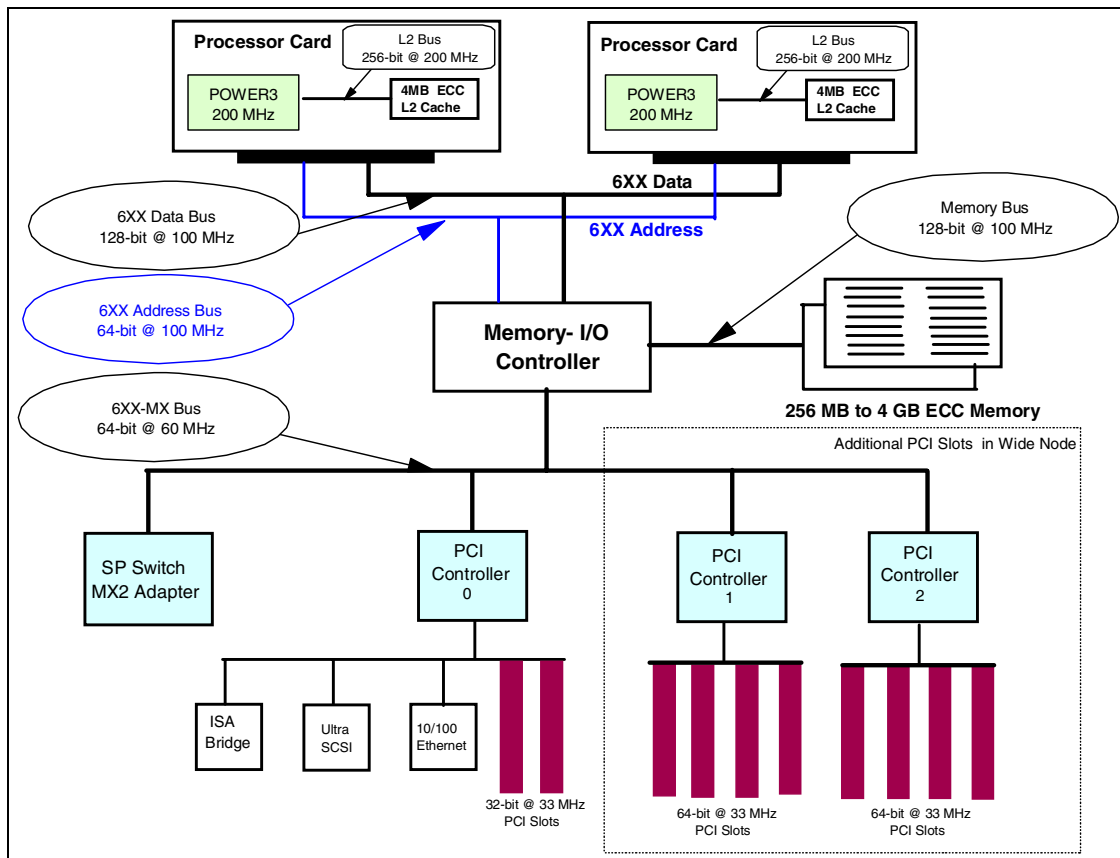


Figure 10. POWER3 SMP node system architecture block diagram

3.3.1.1 POWER3 microprocessor

The POWER3 design contains a superscalar core, which is comprised of eight execution units, supported by a high bandwidth memory interface capable of performing four floating-point operations per clock cycle. The POWER3 design allows concurrent operation of fixed-point, load/store, branch, and floating-point instructions. There is a 32 KB instruction and 64 KB data level 1 cache integrated within a single chip in .25 um CMOS technology. Both instruction and data caches are parity protected. The level 2 cache controller is integrated into the POWER3 microprocessor with the data arrays and directory being implemented with external SRAM modules. The POWER3 microprocessor has a dedicated external interface (separate from 6xx bus interface) for the level 2 cache accesses. Access to the 6xx bus and the level 2 cache can occur simultaneously. The level 2 cache is a unified cache (that is, it holds both instruction and data) and is configured for direct mapped configuration. The external interface to the 4 MB of level 2 cache has 256-bit width and operates at 200 MHz. This interface is ECC protected. The POWER3 microprocessor is designed to provide high performance floating-point computation. There are two floating-point execution units, each supporting 3-cycle latency, 1-cycle throughput, and double/single precision Multiply-Add execution rate. Hence, the POWER3 microprocessor is capable of executing four floating-point operations per clock cycle, which results in a peak throughput of 800 MFLOPS.

3.3.1.2 6xx bus

The 6xx bus or System Bus, as shown in Figure 10 on page 49, connects up to two processor cards to the memory-I/O controller chip set. This bus is optimized for high performance and multiprocessing applications. It provides 40 bits of real address and a separate 128-bit data bus. The address, data, and tag buses are fully parity checked, and each memory or cache request is range checked and positively acknowledged for error detection. Any error will cause a machine check condition and is logged in AIX error logs. The 6xx bus runs at a 100 MHz clock rate, and peak data throughput is 1.6 GB/second. Data and address buses operate independently in true split transaction mode and are pipelined so that new requests may be issued before previous requests are snooped or completed.

3.3.1.3 System memory

The SP POWER3 SMP system supports 256 MB to 4 GB of 10ns SDRAM. System memory is controlled by the memory-I/O chip set via the memory bus. The memory bus consists of a 128-bit data bus and operates at 100 MHz clock cycle. As shown in Figure 10 on page 49, this bus is separated from the System Bus (6xx bus), which allows for concurrent operations on these two buses. For example, cache to cache transfers can occur while a DMA

operation is in progress to an I/O device. There are two memory cards slots in the system. Each memory card contains 16 DIMM slots. Only 128 MB memory DIMMs are supported. Memory DIMMs must be plugged in pairs, and at least one memory card with minimum of 256 MB of memory must be plugged in for system to be operational. System memory is protected by Single Error Correction, Double Error Detection ECC code.

3.3.1.4 I/O subsystem

The memory-I/O controller chip set implements a 64-bit plus parity, multiplexed address and data bus (6xx-MX bus) for attaching three PCI bridge chips and the SP Switch MX2 Adapter. The 6xx-MX bus runs at 60 MHz concurrently and independently from the 6xx and memory buses. At a 60 MHz clock cycle, the peak bandwidth of the 6xx-MX bus is 480 MBps. The three PCI bridge chips attached to 6xx-MX bus provides the interface for 10 PCI slots. Two 32-bit PCI slots are in thin node, and eight additional 64-bit PCI slots are in wide node.

One of the PCI bridge chips (Bridge Chip0) provides support for integrated Ultra2 SCSI and 10Base2, 100BaseT Ethernet functions. The Ultra2 SCSI interface supports up to four internal disks. An ISA bridge chip is also attached to PCI Bridge Chip0 for supporting two serial ports and other internally used functions in the POWER3 SMP node.

3.3.1.5 Service processor

The service processor function is integrated on the I/O planar board in the POWER3 SMP node. Service processor function is for initialization, system error recovery, and diagnostics. The service processor supports system diagnostics by saving the state of the system in a 128 KB non-volatile memory (NVRAM). The service processor code is stored in a 512 KB of flash memory and uses 512 KB of SRAM to execute. The service processor has access to latches and registers on the POWER3 microprocessors and has access to the memory-I/O controller chip set using the serial scan method.

3.3.1.6 System firmware and RTAS

The POWER3 SMP node system firmware flash memory is located on the I/O planar. System firmware contains code that is executed by the POWER3 microprocessor during the initial program load (IPL) phase of the system boot. It also supports various interactions between the AIX operating system and hardware. The extent and method of interaction is defined in the RS/6000 Platform Architecture (RPA). The Run Time Abstraction Software (RTAS), defined by RPA, provides support for AIX and hardware for specific functions, such as initialization, power management, time of day, I/O configuration, and

capture and display of hardware indicators. RTAS and system IPL code are contained on 1 MB of flash memory.

3.3.1.7 System packaging

The POWER3 SMP node system packaging is somewhat different from that of the earlier 332 MHz SMP node. However, the external dimensions and many internal mechanical packaging features have remained the same. One noticeable difference is the absence of flex cables in the POWER3 SMP node. Flex cables were used in 332 MHz nodes to connect the thin node drawer to the expansion I/O drawer. In the POWER3 SMP nodes, connection between the main CPU drawer and expansion I/O drawer is made directly by mating the connectors on I/O planars located in each drawer.

3.3.2 POWER3 SMP High Node (F/C 2054)

POWER3 SMP High Nodes (F/C 2054) use PCI bus architecture and have either two, four, six, or eight 222 MHz 64-bit processors per node. Your IBM RS/6000 SP system must be operating at PSSP 3.1.1 (or later) to use these nodes.

The POWER3 High Node provides additional DASD and PCI adapter capacity by connecting to SP Expansion I/O Units.

The POWER3 SMP High Node occupies two full drawer locations, thus four nodes can be housed in a tall (1.93 m) frame. POWER3 SMP High Nodes can be placed in the first node slot of a frame without requiring additional nodes.

POWER3 SMP High Nodes require a tall, deep frame (Model 550 or F/C 1550); they are supported in the withdrawn 2.01 m frame or in the 1.25 m frame.

Note

POWER3 SMP High Nodes are not compatible with High Performance switches (F/C 4010 and F/C 4007).

3.3.3 PCI bus description

The POWER3 SMP High Node PCI bus contains one 32-bit and four 64-bit PCI slots for I/O adapters.

Additional PCI adapters can be attached to the bus by using up to six optional SP Expansion I/O Units. Each expansion unit has eight 64-bit PCI adapter slots.

3.3.4 Requirements

POWER3 SMP High Nodes occupy two full node drawers. Up to four POWER3 SMP High Nodes may be installed in one tall/deep frame. The mandatory requirements are:

- PSSP 3.1.1 (or later) on the processor node, control workstation, and backup nodes
- Two Processors (on one card, mounted in one slot)
- 1 GB of memory
- 9.1 GB mirrored DASD (with internal booting)

3.3.5 Options

Available options include the following:

- Four processors slots allowing a maximum of eight processors per node
- Four memory slots supporting up to 16 GB of memory
- Five PCI slots (four 64-bit and one 32-bit) for communication adapters
- A dedicated Mezzanine Bus (MX) slot for an optional switch adapter
- Integrated Ethernet with BNC and RJ45 ports (only one port can be used at a time):
 - 10Base2 Ethernet (BNC)
 - 10BaseT or 100BaseTX Ethernet (RJ45)
- Support for up to six SP Expansion I/O Units (F/C 2055)
- Two internal DASD bays supporting up to 72.8 GB of storage (36.4 GB mirrored)
- Integrated Ultra SCSI network
- External nine-pin RS-232 on the planar S2 port (supported only for HACMP serial heartbeat); a 9 to 25-pin converter cable is included with the node
 - Node-to-node HACMP cable (F/C 3124)
 - Frame-to-frame HACMP cable (F/C 3125)

3.3.6 Processor requirements and options

POWER3 SMP High Nodes require a minimum of two 222 MHz, PowerPC processors mounted on one card. You can order up to three additional two-processor cards (F/C 4849) to configure the node with a total of sixteen

CPUs. Table 7 shows the processor options for the POWER3 SMP High Nodes.

Table 7. POWER3 SMP High Node (F/C 2054) processor options

F/C	Quantity	Description	Comments
4849	1 - 4	One processor card with two CPUs	1 required

3.3.6.1 Memory requirements and options

POWER3 SMP High Nodes have one to four memory cards, a minimum of one GB (8 x 128 MB DIMMs) and a maximum of 16 GB (16 x 8 X 128 MB DIMMs). Refer to Table 8 for the memory features for the POWER3 SMP High Nodes.

Table 8. POWER3 SMP High Node (F/C 2054) memory features

F/C	Description	Minimum Node Requirement	Maximum Allowed Per Node
4880	Base memory card	1	4
4133	(8) 128 MB DIMMs	1	16

3.3.7 Disk requirements and options

POWER3 SMP High Nodes can have one pair of internal DASD attached through an integrated Ultra SCSI network. The node can have either no internal DASD (with external booting) or from 9.1 GB to a maximum of 36.4 GB of mirrored, internal disk storage.

Additional DASD can be attached to the POWER3 High Node by using up to six SP Expansion I/O units. Each expansion unit has four DASD bays.

Optional internal disk devices are available as follows:

- 9.1 GB Ultra SCSI disk pair (F/C 2909)
- 18.2 GB Ultra SCSI disk pair (F/C 2918)
- 9.1 GB Ultra SCSI 10K RPM disk pair (F/C 3804)
- 18.2 GB Ultra SCSI 10K RPM disk pair (F/C 3810)
- 36.4 GB Ultra SCSI 10K RPM disk pair (F/C 3820)

External storage devices can be accessed through optional Ultra SCSI adapter (F/C 6207) and SSA adapter (F/C 6230).

3.3.8 Switch and communication adapter requirements and options

If you are planning to use a switch in your SP system, you need switch adapters to connect each RS/6000 SP node to the switch subsystem. SP switch adapters have several characteristics and restrictions, including the following.

Switch restrictions

POWER3 SMP High Nodes are not supported with the SP Switch-8. You must use the SP Switch, 16-port (F/C 4011) or the SP Switch2, 16-port (F/C 4012).

Note

The POWER3 SMP High Node is not compatible with the older High Performance series of switches. If you install a POWER3 Wide Node into a switch-configured system, it must use only SP-type switches.

Switch adapters

The switch adapter for POWER3 SMP High Nodes does not occupy a PCI slot; it is installed into the Mezzanine (MX) bus. The MX bus connects the I/O planar with the system planar. Placing the switch adapter in the MX bus enables switch traffic to proceed at higher bandwidths and lower latencies.

For SP Switch systems, these nodes require the SP Switch MX2 Adapter (F/C 4023).

For SP Switch2 systems, these nodes require the SP Switch2 Adapter (F/C 4025)

I/O adapters

The POWER3 SMP High Node has five PCI (Peripheral Component Interconnect) adapter slots. A full line of PCI adapters is offered for these nodes including:

- SCSI-2
- Ethernet
- Token Ring
- FDDI
- ATM
- Async
- Wide Area Network (WAN)
- SSA RAID5

- S/390 ESCON
- Serial HIPPI

Note

A 10Base2 or 10BaseT/100BaseTX Ethernet adapter for the SP Ethernet is integrated into the POWER3 High Node and does not occupy a PCI slot.

3.3.9 SP Expansion I/O unit (F/C 2055)

Each SP Expansion I/O Unit is an extension of the POWER3 SMP High Node, providing eight additional PCI adapter slots and four DASD bays. PCI adapter hot-plug capability is supported for the SP Expansion I/O Unit with AIX 4.3.3 software loaded on the node.

Up to six expansion units can be connected to each processor node in one to three loops of one or two expansion units in each loop.

Each expansion unit (or pair of units) requires a mounting shelf (F/C 9935). This shelf occupies the space of one drawer in a frame. If only a single expansion unit is mounted in the shelf, a filler plate (F/C 9936) is required for the other side.

Expansion units can be mounted in the same frame as the node, using 2 m cables (F/C 3126), or in separate frames using 15 m cables (F/C 3127). These units require a tall, deep frame (Model 550 or F/C 1550); they are not supported in the withdrawn 2.01 m frame or in the 1.25 m frame.

SP Expansion I/O Unit

IBM suggests that SP Expansion I/O Units be mounted in separate frames, so as not to interfere with switch port utilization.

3.3.10 DASD options

Each SP Expansion I/O Unit has four DASD bays, supporting one or two pairs of DASD.

SCSI and SSA type DASD cannot be mixed within an expansion unit.

Optional DASD pairs for SP Expansion I/O Units are available as follows:

- 9.1 GB Ultra SCSI disk pair (F/C 3800) - requires adapter (F/C 6206)

- 18.2 GB Ultra SCSI disk pair (F/C 3803) - requires adapter (F/C 6206)
- 9.1 GB SSA disk pair (F/C 3802) - requires adapter (F/C 6230)
- 9.1 GB Ultra SCSI 10K RPM disk pair (F/C 3805) - requires adapter (F/C 6206)
- 18.2 GB Ultra SCSI 10K RPM disk pair (F/C 3811) - requires adapter (F/C 6206)
- 18.2 GB SSA disk pair (F/C 3812) - requires adapter (F/C 6230)
- 36.4 GB Ultra SCSI 10K RPM disk pair (F/C 3821) - requires adapter (F/C 6206) and an SP Expansion I/O Unit power upgrade (F/C 9955)
- 36.4 GB SSA 10K RPM disk pair (F/C 3822) - requires adapter (F/C 6230) and an SP Expansion I/O Unit power upgrade (F/C 9955)

Note

Empty, unused DASD bay pairs require a filler plate (F/C 9612).

3.4 332 MHz SMP nodes

On April 21, 1998, IBM first introduced the PCI-based nodes, the RS/6000 SP 332 MHz SMP thin nodes and wide nodes. They provide two or four way symmetric multiprocessing (SMP) utilizing PowerPC technology and extend the RS/6000 PCI I/O technology to the SP system. These PCI-based SMP nodes provide outstanding price/performance in addition to the SP SMP high nodes.

With their outstanding integer performance, 332 MHz PCI SMP nodes are ideal for users who need mission-critical commercial computing solutions to address data mining and data warehouse and online transaction processing (OLTP) applications as well as collaborative computing comprised of Lotus Notes, Domino Server, Internet, intranet, extranet, and groupware application solutions.

The 332 MHz SMP node is also a solid offering for many scientific and technical computing applications as well. As SMP processors in the thin and wide form factors, the 332 MHz nodes provide an ideal way for scientific and technical computing users to explore how to utilize the shared-memory programming model for their applications.

Applications with a mix of fixed-point and floating-point computation, and/or high I/O bandwidth requirements, will also perform well with these SP nodes.

3.4.1 332 MHz SMP node system architecture

The 332 MHz SMP node is designed as a bus-based symmetric multiprocessor (SMP) using a 64-bit address and a 128-bit data system bus running at a 2:1 processor clock ratio. Attached to the system bus are from one to four PowerPC 604e processors with dedicated, in-line L2 cache/bus converter chips and a two chip memory-I/O controller. Note that the 604e processor only uses a 32-bit address bus. The memory-I/O controller generates an independent, separately clocked mezzanine I/O bridge bus to which multiple chips can be attached to implement various architected I/O buses (for example, PCI) that also operate independently and are separately clocked.

This novel design partitions all the system logic into a high speed processor-memory portion and a lower speed I/O portion. This has the cost advantage of not having to design I/O bridges to run wide buses at high speeds, and it also removes electrical loading on the SMP system bus, which allows that bus to run even faster. The wide, high speed processor-memory bus reduces memory and intervention latency, while the separate I/O bridge bus supports memory coherent I/O bridges on a narrower, more cost effective bus. The memory-I/O controller performs all coherency checking for the I/O on the SMP system bus but relieves the SMP system bus from all I/O data traffic.

The 332 MHz SMP node system structure is shown in Figure 11.

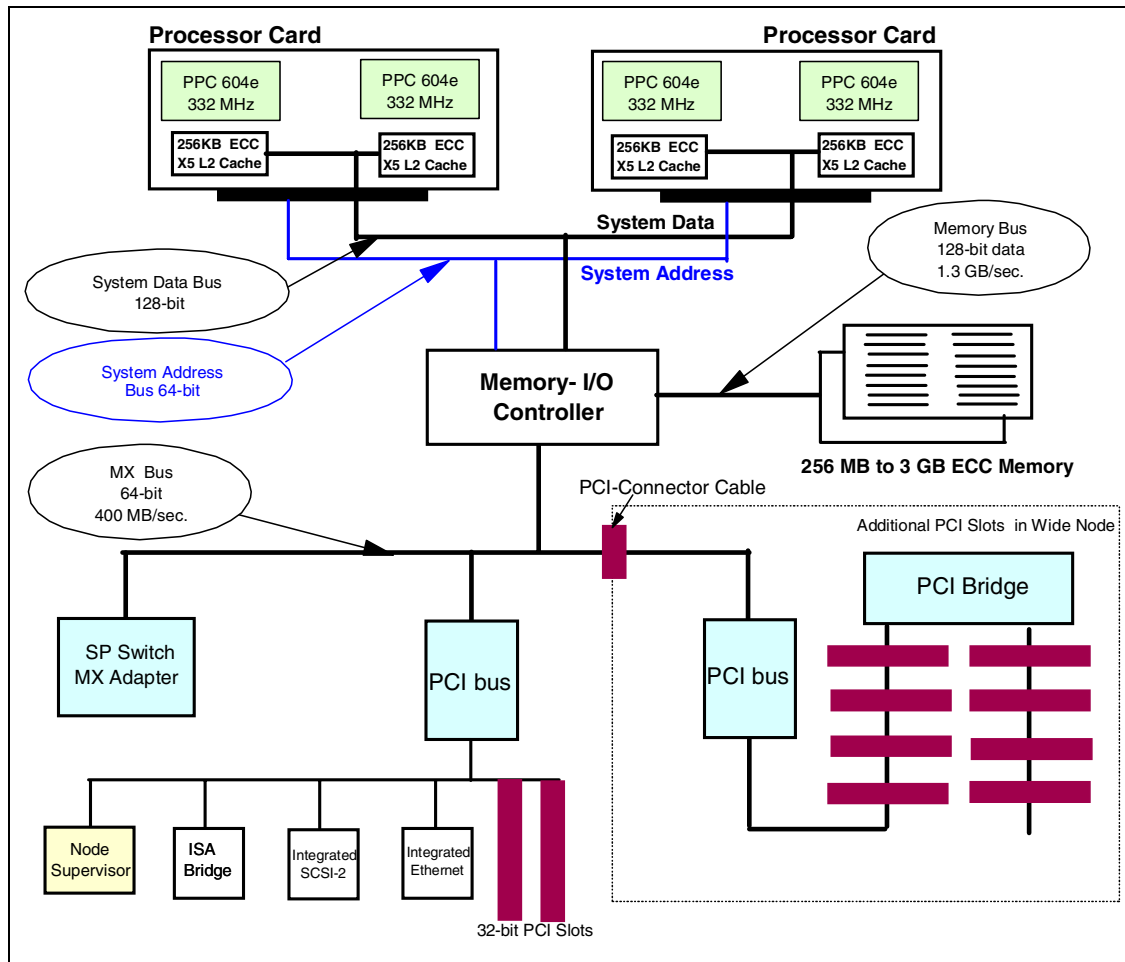


Figure 11. 332 MHz SMP node system architecture block diagram

3.4.1.1 SMP system bus

The system bus of the 332 MHz SMP node is optimized for high performance and multiprocessing applications. It has a separate 64-bit + parity address bus and a 128-bit + parity data bus. These buses operate independently in true split transaction mode and are aggressively pipelined. For example, new requests may be issued before previous requests are snooped or completed. There is no sequential ordering requirement, and each operation is tagged with an 8-bit tag, which allows a maximum of up to 256 transactions to be in progress in the system at any one time. The address bus includes status and

coherency response buses for returning flow control, error reports, or coherency information for each request. It can support a new request every other bus cycle at a sustained rate of over 40 million/sec.

The coherency protocol used is an enhanced Modified Exclusive Shared and Invalid (MESI) protocol that allows for cache-to-cache transfers (intervention) on both modified and shared data found in the L2 or L1 caches. The coherency and data transfer burst size is 64 bytes. The bus was designed for glueless attachment using latch-to-latch protocols and fully buffered devices. This enables high system speeds, and the 332 MHz SMP node can achieve a maximum data rate of 1.3 GB/sec when the data bursts are brick-walled (for example, four quad words are transferred every 4 cycles).

The address, data, and tag buses are fully parity checked, and each memory or cache request is range checked and positively acknowledged for error detection. Any error will cause a machine check condition and be logged.

3.4.1.2 Processor/L2 cache controller

The 332 MHz SMP node X5 Level 2 (L2) Cache controller incorporates several technological advancements in design providing greater performance over traditional cache designs. The Cache controller is inline and totally contained in one chip.

Integrated on the same silicon die as the controller itself are dual tag directories and 256 KB of SRAM cache. The dual directories allow non-blocking access for processor requests and system bus snoops. These directory arrays are fully parity checked, and, in the case of a parity error, the redundant array will be used. The data arrays employ ECC checking and correction, and a single bit error can be detected and corrected with no increase in the cache latency. Multiple bit ECC errors will cause a machine check condition. The cache is configured as an 8-way set-associative, dual sectored, 64 byte line cache. Internal design trade-off performance studies have shown that for commercial workloads the miss rate is comparable to a 1MB direct mapped L2 cache with 32 byte lines.

The L2 cache has two independent buses: one 60x bus connecting with the PowerPC 604e microprocessor and one bus connecting to the SMP system bus. The 60x bus (64 data bits) operates at the same speed as the processor core (166 MHz) for a maximum data bandwidth of 1.2 GB/s. The L2 cache core logic runs 1:1 with the processor clock as well and can source L2 data to the processor in 5:1:1:1 Processor clock cycles (5 for the critical double word, 1 additional clock for each additional double word). The SMP bus interface (128 data bits) operates at a 2:1 ratio (83 MHz) of the processor bus sustaining the 1.3 GB/s maximum data bandwidth.

Since the L2 cache maintains inclusion with the processor's L1 cache, it filters the majority of other processor snoop transactions, thus, reducing the contention for the L1 cache and increasing processor performance.

332 MHz SMP node X5 cache controller supports both the shared and modified intervention protocols of the SMP bus. If one L2 requests a cache line that another L2 already owns, the memory subsystem is bypassed, and the cache line is directly transferred from one cache to the other. The typical latency for a read on the system bus that hits in another L2 cache is 6:1:1:1 bus cycles compared to a best case memory latency of 14:1:1:1 bus cycles. This will measurably reduce the average memory latency when additional L2 caches are added to the system and accounts for the almost linear scalability for commercial workloads.

3.4.1.3 Memory-I/O controller

The 332 MHz SMP node has a high performance memory-I/O controller capable of providing sustained memory bandwidth of over a 1.3 Gigabyte/sec and sustained maximum I/O bandwidth of 400 MBps with multiple active bridges.

The memory controller supports one or two memory cards with up to eight increments of synchronous dynamic RAM (SDRAM) memory on each card. Each increment is a pair of dual in-line memory modules (DIMMs). There are two types of DIMMs supported: 32 MB DIMMs that contain 16 Mb technology chips, or 128 MB DIMMs that contain 64 Mb technology chips. The DIMMs must be plugged in pairs because each DIMM provides 72 bits of data (64 data bits + 8 bits of ECC word), which, when added together form the 144 bit memory interface. The memory DIMMs used in the 332 MHz SMP node are 100 MHz (10 ns), JEDEC standard non-buffered SDRAMs.

SDRAMs operate differently than regular DRAM or EDO DRAM. There is no interleaving of DIMMs required to achieve high data transfer rates. The SDRAM DIMM can supply data every 10 ns clock cycle once its access time is satisfied, which is 60 ns. However, as with a traditional DRAM, that memory bank is then busy (precharging) for a period before it can be accessed again. So, for maximum performance, another bank needs to be accessed in the meantime. The memory controller employs heavy queuing to be able to make intelligent decisions about which memory bank to select next. It will queue up to eight `read` commands and eight `write` commands and attempt to schedule all reads to non-busy banks and fill in with writes when no reads are pending. Writes will assume priority if the write queue is full or there is a new read address matching a write in the write queue.

The SDRAM DIMM itself contains multiple internal banks. The 32 MB DIMM has two internal banks, and the 128 MB DIMM has four internal banks. A different non-busy bank can be accessed to keep the memory fully utilized while previously accessed banks are precharging. In general, the more banks of memory the better to allow the controller to schedule non-busy banks. So, 32 MB DIMMs with 2 banks internal will contain more banks than the same amount of memory in 128 MB DIMMs. For example, suppose you wanted 256 MB of memory total. That would be 1 pair of 128 MB DIMMs with a total of 4 internal banks or 4 pairs of 32 MB DIMMs with 2 internal banks each for a total of 8 banks. Once larger amounts of memory are installed (for example, more than 16 banks), it should not make much difference which DIMMs to use for there would be plenty of non-busy banks to access. Note that there is no performance advantage to spreading DIMMs across two memory cards. The second card is needed only for expanding beyond a full first card.

Assuming the read requests can be satisfied to non-busy banks, the memory controller is capable of supplying a quadword of data every 83 MHz clock for an aggregate peak bandwidth of 1.32 GB/sec. While the memory bandwidth must be shared with any I/O, the SMP data bus bandwidth does not. This means that concurrent memory or cache-to-cache data can be sent on the SMP data bus while the memory controller is using the I/O bridge bus to transfer data or from I/O. There is also extensive queuing to support the I/O bus, seven command queues to hold outgoing processor requests to I/O, and six command queue's forwarding I/O requests for Direct Memory Access (DMA).

To provide memory coherency for I/O but allow the SMP bus to run at maximum speed, the memory-I/O controller acts as a snooping agent. This is accomplished by keeping track of the latest I/O reads in a chip directory. This directory is used whenever a processor request is issued on the SMP bus so that the request does not have to be forwarded to I/O to be snooped. When I/O issues a request, the controller places the request address on the SMP bus to be snooped by the processor caches before it is entered into the memory queues. A cache hit will retry the request and force the data to memory and then the request will be reissued by the memory controller on behalf of the I/O device. It is not rejected back to the device.

3.4.1.4 I/O bridge bus

The 332 MHz SMP node memory-I/O controller implements a 64-bit + parity, multiplexed address and data bus for attaching several PCI I/O bridges. This bus runs concurrent and independently from the SMP processor bus and the SDRAM memory buses. In the 332 MHz SMP node, this I/O bridge bus is clocked at 50 MHz, and there are three PCI bridges attached. The peak

bandwidth of this bus is 400 MBps, and it is also a split-transaction, fully tagged bus, like the SMP system bus. Multiple requests from each I/O bridge and the processors may be active at any one time, therefore, subject only to the amount of queuing and tags available in the bridges and the memory-I/O controller.

This bus is parity checked, and all addresses are range checked and positively acknowledged. In addition, the memory-I/O controller and the I/O bridges have fault isolation logic to record any detected error so that software can more easily determine the source of an error.

3.4.1.5 PCI I/O bridges

Each PCI hardware bridge chip attaches to the I/O bridge bus and is separately and concurrently accessible by any of the processors. There are no PCI bridge-bridge chips. All three of the 332 MHz SMP node's PCI buses have their own bridge chip designed to the PCI version 2.0 Local Bus Specification. These PCI bridges contain over 700 bytes of internal data buffering and will support either a 32-bit or 64-bit PCI data bus configuration. The internal data buffer structure of the bridges allows the I/O bridge bus interface functions and the PCI interface functions to operate independently. Processor initiated Load/ Store operations may be up to 64 bytes long and are buffered separately from PCI adapter initiated DMA operations. Consequently, acceptance of processor Load/Store operations that target the PCI bus only depend upon buffer availability and not the availability of the target bus or completion of DMA operations. Likewise, the bridge will always accept PCI bus initiator DMA read or write operations that target system memory sends if there is an available data buffer and queue available, and it does not have to wait for previous reads or writes to be completed. These PCI addresses are translated by the bridge chip using software initialized translation control elements (TCEs) fetched from memory. Each TCE will translate a 4K page of PCI address space.

The bridge design is optimized for large block sequential DMA read and DMA write operations by a PCI Bus Initiator. Up to 128 bytes of DMA read data are continuously prefetched into one of four DMA read caches in order to sustain large block transfers. Two outstanding DMA read requests may be active at one time by each PCI bridge. When the transfer is terminated, the data and its TCE remain in the cache until the cache is reassigned or invalidated by a processor write to memory at that cached location. Each bridge also maintains an eight-entry TCE cache for PCI initiator DMA write operations and a 128 byte write-thru DMA buffer. The bridge PCI and system interfaces are parity checked and contain error detection and error reporting registers accessible by software for fault isolation. The combination of high performing,

independent I/O bridges and the node split bus architecture yields outstanding I/O performance and scalability.

3.4.2 332 MHz SMP thin nodes (F/C 2050)

332 MHz SMP thin nodes (F/C 2050) have PCI bus architecture and use either two or four 332 MHz PowerPC processors per node. These nodes are functionally equivalent to an IBM RS/6000 7025-F50 workstation. The SP system must be operating at PSSP 2.4 (or later) to use these nodes.

The 332 MHz SMP thin node occupies one half of a drawer and may be installed singly with systems operating at PSSP 3.1 or later. Therefore, up to 16 SMP thin nodes can be housed in a tall frame. When installed singly, 332 MHz SMP thin nodes must be placed in the odd numbered node slot. See 3.4.2.2, "Single SMP thin node configuration rules" on page 64, for details.

For electromagnetic compliance, these nodes are housed in an SMP Enclosure. This enclosure (F/C 9930) is automatically included when you order a 332 MHz SMP thin node. For installations using single SMP thin nodes, a cover plate (F/C 9931) is also included to cover the unused enclosure slot.

If you are going to mount a 332 MHz SMP thin node into an older 2.01 m or 1.25 m frame, a power system upgrade is necessary. However, once you have done the power system upgrade, these nodes are fully compatible with all existing SP hardware except for High Performance Switches.

3.4.2.1 Bus description

The 332 MHz SMP thin node PCI bus contains two 32-bit slots PCI slots (slots I2 and I3). The I1 slot is reserved for the optional SP Switch MX2 Adapter. Previously installed 332 MHz SMP thin nodes may have a withdrawn SP Switch MX Adapter in the I1 slot.

3.4.2.2 Single SMP thin node configuration rules

With PSSP 3.1, single POWER3 SMP thin nodes and single 332 MHz. SMP thin nodes are allowed in both tall and short frame configurations provided the following rules are observed:

- Single SMP thin nodes must be installed in the odd numbered node position. Single SMP thin nodes are not supported in the even numbered node position.
- Empty node drawers are allowed on tall frames if the frame is either a non-switched frame or configured with an SP Switch (16-port switch).

- Tall frames configured with the SP Switch-8 (8-port switch) must have all nodes placed in sequential order; no empty drawers are allowed. Therefore, the single SMP thin node in these frames is the last node in the configuration.
- Short frame configurations must have all nodes placed in sequential order; no empty drawers are allowed. Therefore, the single SMP thin node in these frames is the last node in the configuration.
- A single POWER3 SMP thin node and a single 332 MHz SMP thin node each occupy one half of a node drawer.
- Single POWER3 SMP thin nodes and single 332 MHz SMP thin nodes may be mixed in a thin node drawer.
- If a frame has more than six single SMP thin nodes installed, that frame will have an uneven weight distribution. You must be careful when moving these frames.

3.4.2.3 Requirements

332 MHz SMP thin nodes occupy one half of a node drawer. When two SMP thin nodes are placed in one drawer, the nodes may be asymmetrically configured for memory, disk, processor speed, and adapters. Up to sixteen 332 MHz SMP thin nodes may be installed in one tall frame and up to eight in a short frame.

The mandatory requirements are:

- PSSP 2.4 (or later) on the control workstation, backup nodes, and processor node
- Two processors (mounted in one slot)
- 256 MB of memory
- 4.5 GB of DASD (with internal booting)
- An upgraded power system on older frames

3.4.2.4 Options

Each 332 MHz SMP thin node is functionally equivalent to an RS/6000 7025-F50 and has the following options:

- Two processor slots allowing a maximum of four processors per node
- Two memory slots supporting up to 3 GB of memory
- Two disk bays supporting up to 36.4 GB of storage (18.2 GB mirrored)
- A dedicated Mezzanine Bus (MX) slot for an optional switch adapter

- Two 32-bit PCI slots for communication adapters
- Integrated 10BaseT/10Base2 Ethernet (only one port may be used at one time).
- Integrated SCSI-2 Fast/Wide.
- Standard Service Processor.
- External nine-pin RS-232 on the planar S2 port.
 - This connection has active heartbeat and is available for customer applications.

3.4.2.5 Processor requirements and options

SMP thin nodes require a minimum of two 332 MHz PowerPC processors mounted on one card. However, you can order an additional processor card (F/C 4320) to configure the node with a total of four CPUs.

Table 9 provides the processor options for the 332 MHz SMP thin nodes.

Table 9. 332 MHz SMP Thin Node processor options

F/C	Multiplier	Description	Comments
4320	X 1	One processor card with two CPUs	Minimum required
4320	X 2	Two processor cards with two CPUs each (four CPUs total)	Maximum allowed

3.4.2.6 Memory requirements and options

332 MHz SMP thin nodes have two memory cards and require a minimum of 256 MB of memory. These nodes support a maximum of 3 GB of memory. Memory is supplied by 128 MB DIMMs that must be mounted in pairs (256 MB increments). The memory cards are not required to be configured symmetrically. Each card has the capacity to mount 2 GB of DIMMs; however, only 3 GB are addressable per node. Memory cards and DIMMs are not interchangeable between SMP and non-SMP thin nodes.

Table 10 provides the available memory features for the 332 MHz SMP thin nodes.

Table 10. 332 MHz SMP Thin Node memory features

F/C	Description	Minimum Node Requirement	Maximum Allowed Per Node
4093	Base Memory Card	2	2
4110	One Pair of 128 MB DIMMs (256 MB total)	One pair	Six pairs

3.4.2.7 Disk requirements and options

332 MHz SMP thin nodes can have up to two internal disks attached through an integrated SCSI-2 network. The 332 MHz SMP thin node can have either no internal disk (with external booting) or from 4.5 GB to a maximum of 36.4 GB of internal disk storage. External storage devices can be accessed through an optional Ultra SCSI Adapter (F/C 6207) or SCSI-2 Adapter (F/C 6209).

Optional direct access storage devices are available as follows:

- 4.5 GB Ultra SCSI disk drive (F/C 2900)
- 4.5 GB Ultra SCSI disk drive pair (F/C 2904)
- 9.1 GB Ultra SCSI disk drive (F/C 2908)
- 9.1 GB Ultra SCSI disk drive pair (F/C 2909)
- 18.2 GB Ultra SCSI disk drive pair (F/C 2918)

This node does not require special cables or adapters to mount internal disk.

3.4.2.8 Switch adapter requirements and options

The switch adapter for SMP thin nodes does not occupy a PCI slot. Instead, the switch adapter for these nodes is installed into the Mezzanine (MX) bus. The MX bus connects the I/O planar with the system planar. Placing the switch adapter in the MX bus enables switch traffic to proceed at higher bandwidths and lower latencies.

In switch configured systems, 332 MHz SMP thin nodes require the following switch adapter:

- SP Switch MX2 Adapter (F/C 4023)

332 MHz SMP thin node switch restrictions

The 332 MHz SMP thin node is not compatible with the older High Performance series of switches. If an SMP thin node is going to be placed into an SP system configured with a switch, that switch must be either an SP Switch or an SPS-8 switch.

Switch adapters for SMP thin nodes are not interchangeable with switch adapters used on 160 MHz thin node.

3.4.3 332 MHz SMP wide nodes (F/C 2051)

332 MHz SMP wide nodes (F/C 2051) have PCI bus architecture and use either two or four 332 MHz PowerPC processors per node. These nodes are

functionally equivalent to an IBM RS/6000 7025-F50 workstation. The SP system must be operating at PSSP 2.4 (or later) to use these nodes.

The 332 MHz SMP wide node occupies one full drawer; therefore, eight SMP wide nodes can be housed in a tall frame. SMP wide nodes can be placed in the first node slot of a frame without requiring additional nodes.

For electromagnetic compliance, these nodes are housed in an SMP enclosure. This enclosure (F/C 9930) is automatically included when you order a 332 MHz SMP wide node.

If you are going to mount a 332 MHz SMP wide node into an older 2.01 m or 1.25 m frame, a power system upgrade is necessary. However, once you have done the power system upgrade, these nodes are fully compatible with all existing SP hardware except for High Performance Switches.

3.4.3.1 Bus description

The 332 MHz SMP wide node PCI bus is divided into three logical groups of PCI slots. The first slot group (slots I2 and I3) is composed of the two 32-bit slots residing on the CPU side of the 332 MHz SMP wide node, and the second and third group reside on the I/O side of the node. Both the second and third group have four PCI slots each. The second group (slots I1 through I4) has three 64-bit slots and a single 32-bit slot. The third group (slots I5 through I8) is made up of the last four 32-bit slots on the I/O side of the node. The third group is a physical extension on the second group. The I1 slot on the CPU side of the node is reserved for the optional SP Switch MX2 Adapter. Previously installed 332 MHz SMP wide nodes may have a withdrawn SP Switch MX Adapter in the CPU side I1 slot.

Adapter placement restrictions

With few exceptions, the ten PCI slots in the 332 MHz SMP wide node can be used for any valid RS/6000 SP PCI system adapter. While most PCI adapters will function in any 332 MHz SMP wide node slot, the following adapters cannot be placed in any one of the third group of PCI slots:

- S/390 ESCON (F/C 2751)
- ARTIC960Hx 4-Port selectable (F/C 2947)
- 2-port Multiprotocol (F/C 2962)
- ATM 155 UTP (F/C 2963)
- Gigabit Ethernet - SX (F/C 2969)
- ATM 155 MMF (F/C 2988)
- Ultra SCSI SE (F/C 6206)

- Ultra SCSI DE (F/C 6207)
- SSA RAID5 (F/C 6215)
- ARTIC960RxD Quad Digital Trunk (F/C 6310)

To achieve the best performance with SSA RAID and Ultra SCSI disk subsystems, the following adapters for these devices should be distributed evenly across the two recommended PCI slot groups:

- SSA RAID5 (F/C 6215)
- Ultra SCSI SE (F/C 6206)
- Ultra SCSI DE (F/C 6207)

To avoid performance degradation, the following adapters should not be placed in slots I5, I6, I7, or I8 in 332 MHz SMP wide nodes:

- FDDI SK-NET LP SAS (F/C 2741)
- FDDI SK-NET LP DAS (F/C 2742)
- FDDI SK-NET UP SAS (F/C 2743)
- 10/100 MB Ethernet (F/C 2968)
- SCSI-2 F/W single-ended (F/C 6208)
- SCSI-2 F/W differential (F/C 6209)

For similar reasons, if two S/390 ESCON adapters (F/C 2751) are used in this node, one adapter must be placed in the CPU bus, and the other adapter must be placed in the first I/O bus.

3.4.3.2 Requirements

332 MHz SMP wide nodes occupy one full node drawer. These nodes are symmetrically configured for memory, disk, and adapters. Up to eight 332 MHz SMP wide nodes may be installed in one tall frame and up to four in a short frame. The mandatory requirements are:

- PSSP 2.4 (or later) on the control workstation, backup nodes, and processor node
- Two processors (mounted in one slot)
- 256 MB of memory
- 4.5 GB of DASD (with internal booting)
- An upgraded power system on older frames

3.4.3.3 Options

Each 332 MHz SMP wide node is functionally equivalent to an RS/6000 7025-F50 and has the following options:

- Two processor slots allowing a maximum of four processors per node
- Two memory slots supporting up to 3 GB of memory
- Ten PCI slots for communication adapters (seven 32-bit and 3 64-bit)
- A dedicated Mezzanine Bus (MX) slot for an optional switch adapter
- Integrated 10BaseT/10Base2 Ethernet (only one port may be used at one time)
- Four disk bays supporting up to 72.8 GB of disk storage (36.4 GB mirrored)
- Integrated SCSI-2 Fast/Wide
- Standard Service Processor
- External nine-pin RS-232 on the planar S2 port
 - This connection has active heartbeat and is available for customer applications.

3.4.3.4 Processor requirements and options

SMP wide nodes require a minimum of two 332 MHz PowerPC processors mounted on one card. However, you can order an additional processor card (F/C 4320) to configure the node with a total of four CPUs.

Table 11 provides the processor options for the 332 MHz SMP wide nodes.

Table 11. 332 MHz SMP Wide Node processor options

F/C	Multiplier	Description	Comments
4320	X 1	One processor card with two CPUs	Minimum required
4320	X 2	Two processor cards with two CPUs each (four CPUs total)	Maximum allowed

3.4.3.5 Memory requirements and options

332 MHz SMP wide nodes have two memory cards and require a minimum of 256 MB of memory. These nodes support a maximum of 3 GB of memory. Memory is supplied by 128 MB DIMMs that must be mounted in pairs (256 MB increments).

The memory cards are not required to be configured symmetrically. Each card has the capacity to mount 2 GB of DIMMs; however, only 3 GB are

addressable per node. Memory cards and DIMMs are not interchangeable between SMP and non-SMP wide nodes.

Table 12 provides the available memory features for the SMP wide nodes.

Table 12. 332 MHz SMP Wide Node memory features

F/C	Description	Minimum Node Requirement	Maximum Allowed Per Node
4093	Base Memory Card	2	2
4110	One Pair of 128 MB DIMMs (256 MB total)	One pair	Six pairs

3.4.3.6 Disk requirements and options

332 MHz SMP wide nodes can have up to four internal disks attached through an integrated SCSI-2 network. The 332 MHz SMP wide node can have either no internal disk (with external booting) or from 4.5 GB to a maximum of 72.8 GB of internal disk storage. External storage devices can be accessed through an optional Ultra SCSI Adapter (F/C 6207) or SCSI-2 Adapter (F/C 6209).

Optional direct access storage devices are available as follows:

- 4.5 GB Ultra SCSI disk drive (F/C 2900)
- 4.5 GB Ultra SCSI disk drive pair (F/C 2904)
- 9.1 GB Ultra SCSI disk drive (F/C 2908)
- 9.1 GB Ultra SCSI disk drive pair (F/C 2909)
- 18.2 GB Ultra SCSI disk drive pair (F/C 2918)

This node does not require special cables or adapters to mount internal disks. However, the 332 MHz SMP wide node has an option (F/C 1241) that provides an independent SCSI hookup. It accomplishes the following:

- Eliminates the DASD controller as a single point of failure during mirroring
- Increases disk performance
- Balances disk loading

The (F/C 1241) option requires either an (F/C 6206) SCSI-2 Ultra/Wide Adapter PCI or an (F/C 6208) SCSI-2 Fast/Wide Adapter 4-A PCI as a PCI-type SCSI adapter.

3.4.3.7 Switch adapter requirements and options

The switch adapter for SMP wide nodes does not occupy a PCI slot. Instead, the switch adapter for these nodes is installed into the Mezzanine (MX) bus. The MX bus connects the I/O planar with the system planar. Placing the switch adapter in the MX bus enables switch traffic to proceed at higher bandwidths and lower latencies.

In switch-configured systems, 332 MHz SMP wide nodes require the following switch adapter:

- SP Switch MX2 Adapter (F/C 4023)

332 MHz SMP wide node switch restrictions

The 332 MHz SMP wide node is not compatible with the older High Performance series of switches. If an SMP wide node is going to be placed into an SP system configured with a switch, that switch must be either an SP Switch or an SP Switch-8.

Switch adapters for SMP wide nodes are not interchangeable with switch adapters used on 160 MHz thin node.

Chapter 4. SP-attached servers

The SP-attached server is an IBM RS/6000 7017 Enterprise Server configured to operate with an RS/6000 SP System. The SP configuration requires:

- Use of an SP supported Ethernet card for connection to the SP Ethernet.
- A custom RS-232 cable connecting the SP's control workstation to the 7017's SAMI port.
- A second custom RS-232 cable connecting the SP's control workstation to the 7017's S1 serial port.
- If the SP System is switch configured, the 7017 must also have an optional RS/6000 SP System Attachment adapter (RS/6000 F/C 8396) installed. This adapter uses an SP Switch cable to connect to a valid switch port on an SP Switch.

With this configuration, the SP-attached server enhances the SP System's performance and provides the scalability needed for e-business applications.

Figure 12 shows IBM RS/6000 7017 Enterprise Server.



Figure 12. IBM RS/6000 7017 Enterprise Server

4.1 Overview

Both the 7017-S70 Enterprise Server, the 7017-S7A Enterprise Server (also known as the 7017-S70 Advanced Enterprise Server) and the 7017-S80 Enterprise Server use RS/6000 feature codes not RS/6000 SP feature codes.

Your IBM representative will be able to furnish any feature codes needed to order the options listed in this overview.

The RS/6000 SP Feature Codes associated with the SP-attached server (F/C 9122 plus F/C 9123) refer to the system connections that attach the RS/6000 Enterprise Server to your RS/6000 SP system. F/C 9122 and F/C 9123 do not refer to hardware components. Because the Enterprise Servers are stand-alone devices with cable attachments to the SP system, these servers have some attributes that appear node-like to the SP system and other attributes that appear frame-like.

F/C 9122 Refers to the node-like attachment between the SP-attached server and your SP system.

F/C 9123 Refers to the frame-like attachment between the SP-attached server and your SP system.

Both the 7017-S70, the 7017-S70 Advanced Server, and the 7017-S80 Servers appear nearly identical to your SP system, and F/C 9122 plus F/C 9123 can be used to attach either server. Your IBM representative will help you decide which 7017 Enterprise Server matches your e-business needs.

4.1.1 How the SP system views the SP-attached server

With a few hardware control exceptions, the SP-attached server performs the same functions that standard SP processor nodes perform. However, since the SP-attached server is mounted in its own frame and not in an SP frame, the SP system cannot view the SP-attached server as just another node. Instead, the SP system views the SP-attached server as an object with both frame and node characteristics. The node-like features of the SP-attached server are driven by F/C 9122, while the frame-like features of this device are driven by F/C 9123.

Because the SP-attached server has both frame and node characteristics, it must have both a frame number and a node number. However, since the SP-attached server does not have full SP frame characteristics, it cannot be considered as a standard SP expansion frame. Therefore, when assigning the SP-attached server's frame number, you have to follow two rules:

- The SP-attached server cannot be inserted between a switch configured frame and any non-switched expansion frame using that switch.

For example, frames one and five of an SP system are switch configured. Frame two is a non-switched expansion frame attached to frame one. Frame six, seven, and eight are non-switched expansion frames attached to frame five.

In this configuration, an SP-attached server could be given frame number three, but that would forbid any future attachment of non-switched expansion frames to frame one's switch.

If you assigned the SP-attached server frame number nine, your system could still be scaled using other switch configured frames and non-switched expansion frames. The SP-attached server can be inserted between two switch configured frames.

Once the frame number has been assigned, the server's node number, which is based on the frame number, is automatically generated. The following system defaults are used:

- The SP-attached server is viewed by the system as a single frame containing a single node.
- The system places the server's node-like features in the slot one position.
- Each SP-attached server installed in an SP system subtracts one node from the total node count allowed in the system. However, because the SP-attached server has frame-like features, it reserves 16 node numbers that are used in determining the node number of nodes placed after the attached server.

Server attachment limitations

When you attach a 7017 Enterprise Server to an SP system, certain limitations apply.

- The first frame in the SP system must be an SP frame containing at least one node.
- You can attach up to eight 7017 Enterprise Servers onto an RS/6000 SP system.
- Each SP-attached server requires one valid, unused node slot in the SP system for switch port assignment.
 - An assigned switch port is required in both switch configured and non-switched SP systems.
 - See "Assigning a frame number" on page 81 and "Assigning a switch port number" on page 82 for information on configuring the SP-attached servers.
- In some cases, the number of SP-attached servers you are planning to install may exceed the number of available node slots in an SP frame. If this happens, you can take advantage of any valid, unused node slots (and the associated switch ports) that may exist in other SP frames in your SP system.

For example, you have a two-frame SP system. The first SP frame contains ten thin nodes and an SP Switch. The second SP frame contains five single SMP thin nodes and another SP Switch. You want to attach eight 7017 Enterprise Servers.

In this example, you can attach six of the Enterprise Servers to the first frame and two Enterprise Servers to the second SP frame. As an alternative, all eight SP-attached servers could be connected to the second SP frame.

- In some cases, the number of SP-attached servers you are planning to install may exceed the number of available node slots in your SP system. If this happens, you will need to add an additional SP frame to your SP system.
 - Only the first SP frame is required to have nodes, additional SP frames may be empty.
- Each SP-attached server counts as one node that must be subtracted from the total node count of 128 allowed in an SP system (without special order).
- Each SP-attached server also counts as one frame that must be subtracted from the total frame count allowed in an SP system.

4.2 Installation requirements

There are several requirements for hardware and software that must be met before you can place the SP-attached server into service with your SP system. These requirements are in the following categories:

- System requirements
- Switch adapter requirements
- Network media card requirements
- Software requirements

4.2.1 System requirements

The following hardware requirements must be met before you can place the SP-attached server into service:

- Your SP system must be operating with a minimum of PSSP 3.1 (or later) and AIX 4.3.2 (or later).
 - Each SP-attached server also requires its own PSSP license.

- See Section 4.2.4, “Software requirements” on page 81 for details and system configuration requirements.
- Your SP system must be a tall frame system. Short frames are not compatible with the SP-attached server.
- The system may be switched or non-switched.
- If it is a switched system, the switches must be 16-port SP Switches (F/C 4011). The SP Switch-8 is not compatible with the SP-attached server.
- If the SP System is switch configured, the SP-attached server must also have a special switch adapter installed. See Section 4.2.2, “RS/6000 SP System Attachment adapter (RS/6000 F/C 8396)” on page 77 for adapter details and placement restrictions.
- Three control workstation connections are required:
 1. The SP Ethernet connection from the control workstation must use an SP supported Ethernet card mounted in the SP-attached server.
 2. A custom RS-232 cable must connect the SP system’s control workstation to the SP-attached server’s SAMI port.
 3. A second custom RS-232 cable must connect the SP system’s control workstation to the SP-attached server’s S1 serial port.

For information on these connections, see Section 4.3.1, “Connecting to the control workstation” on page 86 for details and adapter restrictions.

- To ensure that the entire SP system is at the same electrical potential, the frame-to-frame ground cables provided with your SP system must be used between the SP system and the SP-attached server.
- Some cables used with your SP-attached server have limited lengths. You must keep those lengths and any required cable drops in mind when locating your SP-attached server in relation to other SP equipment.

4.2.2 RS/6000 SP System Attachment adapter (RS/6000 F/C 8396)

If you are placing an SP-attached server into a system that uses SP Switches, you must install an SP Switch adapter in the SP-attached server. However, unlike the SP Switch Router, which can have several router-specific switch adapters installed, the SP-attached server is only allowed to have one server-specific switch adapter installed for each server system. The adapter used to connect the SP-attached server to the SP Switch is called the RS/6000 SP System Attachment Adapter (RS/6000 F/C 8396). Because the RS/6000 SP System Attachment adapter is a non-SP adapter, it is ordered using an RS/6000 feature code.

The single RS/6000 SP System Attachment adapter that you place into each SP-attached server requires the following:

- One valid, unused switch port on the SP Switch corresponding to a legitimate node slot in your SP configuration.

A legitimate node slot may be empty, the second half of a wide node, or one of the last three positions of a high node, provided that node slot satisfies the other rules for configuring nodes in an SP system. For example, if you have a frame with 16 thin nodes installed, you cannot attach an RS/6000 SP System Attachment adapter to that frame until you remove a node and delete its configuration from the system image.
- One media card slot (slot 10) in the primary (first) I/O tower of SP-attached server.

See Section 4.2.2.1, “Placement restrictions” on page 78, for other limitations

4.2.2.1 Placement restrictions

The RS/6000 SP System Attachment adapter has the following placement restrictions:

- The RS/6000 SP System Attachment adapter must be installed in slot 10 of the SP-attached server’s I/O tower.
- Slot 9 must be left open to ensure the adapter has sufficient bandwidth.
- Slot 11 must be left open to provide clearance for the switch adapter’s heat sinks.

Installing in existing 7017 Enterprise Servers

If you are attaching an existing 7017 Enterprise Server to an SP System, you may find a SCSI adapter installed in slot 9 of the server. This SCSI adapter must be relocated. However, the SCSI adapter in slot 9 is typically connected to a boot device and requires special attention before removal. The following is the boot device SCSI adapter relocation overview:

1. Boot up the Enterprise Server that you are going to attach to the SP system.
2. Follow standard AIX procedures to change the boot device.
 - Change the device codes.
 - Change the device address.
3. Take the Enterprise Server down.

4. Move SCSI card from slot 9 to the new location (remember, slots 9 and 11 must be left open, and the SP System Attachment adapter must be placed in slot 10)
 - SCSI adapter F/C 6206 and F/C 6208 must be placed in either slot 12 or slot 14.
 - SCSI adapter F/C 6207 and F/C 6209 must be placed in either slot 12, slot 13, or slot 14.
5. Reboot server and continue with SP attachment.

Consult the appropriate documentation for specific installation procedures.

4.2.2.2 Cables

When the RS/6000 SP System Attachment Adapter is ordered, you will also need to order the following cable:

- A 10 m switch cable (F/C 9310)
Connects the RS/6000 SP System Attachment Adapter to a valid switch port on the SP Switch.

There are no optional cables for the RS/6000 SP System Attachment Adapter.

Although the SP System Attachment Adapter is ordered with the Enterprise Server, the 10 m cable (F/C 9310) must be ordered with the SP system.

4.2.3 Network media card requirements

Each network media card requires one media card slot in the SP-attached server. All network adapters in the SP-attached server use PCI architecture.

SP supported PCI adapters only

Only SP supported PCI adapters may be used in a 7017 Enterprise Server when it is used as an SP-attached server. This means that if you are attaching an existing RS/6000 7017 Enterprise Server to an SP system, you must remove any non-SP supported PCI adapters.

Note the rules that govern supported PCI adapters, such as:

- Required adapters (including minimum requirements)
- Maximum number allowed for each adapter
- Bus placement restrictions found in the RS/6000 7017 Enterprise Server documentation

The following adapters, as shown in Table 13, are supported on an RS/6000 Enterprise Server when used as an SP-attached server.

Table 13. Supported communication adapters for SP-attached servers

F/C	PCI adapter name
2741	FDDI SK-NET LP SAS
2742	FDDI SK-NET LP DAS
2743	FDDI SK-NET UP SAS
2751	S/390 ESCON Channel Adapter
2920	Token Ring Auto Lanstream
2943	EIA 232/RS-422 8-port Asynchronous Adapter
2944	WAN RS232 128-port
2947	IBM ARTIC960Hx 4-Port Selectable Adapter
2962	2-port Multiprotocol X.25 Adapter
2963	ATM 155 TURBOWAYS UTP Adapter
2968	Ethernet 10/100 MB
2969	Gigabit Ethernet - SX
2985	Ethernet 10 MB BNC
2987	Ethernet 10 MB AUI
2988	ATM 155 MMF
6206	Ultra SCSI Single Ended
6207	Ultra SCSI Differential
6208	SCSI-2 F/W Single-Ended
6209	SCSI-2 F/W Differential
6215	SSA RAID 5 (accepts optional SSA Fast-Write Cache module (F/C 6222))
6310	IBM ARTIC960RxD Quad Digital Trunk Adapter

With the exception of the RS/6000 SP System Attachment adapter (which is supported but not listed here), if an adapter does not appear in this list, it is not supported in the SP-attached server. If you are planning to use an existing Enterprise Server, and any of the previously installed adapters do not

appear in this list, they must be removed from the Enterprise Server before it can be attached to the SP.

For more information on each of these adapters, see Chapter 9, “Communication adapters” on page 137.

Note

F/C 2985 and 2987 have placement restrictions when used as the SP Ethernet adapter. See Section 4.3.1.1, “Attaching the SP Ethernet” on page 86 for more information.

4.2.4 Software requirements

The SP-attached server requires an SP system operating with:

- PSSP 3.1 (or later)
- AIX 4.3.2 (or later)
- System partitions using lower levels of software are permitted

Note that if you are attaching an existing Enterprise Server, and that server is connected to an IPv6 network, you will need to remove the server from the network before making the SP attachment.

Each SP-attached server also requires its own PSSP license. PSSP is available through:

- F/C 5800 (delivered on 4 mm tape)
- F/C 5801 (delivered on 8 mm tape)
- F/C 5802 (delivered on CD-ROM)

Note that the High Performance series of switches cannot be used with an SP-attached server since these switches are not supported on PSSP 3.1.

4.2.4.1 Software configuration requirements

The SP-attached server requires two software inputs for network configuration:

1. Assigning a frame number
2. Assigning a switch port number

Assigning a frame number

SP-attached servers are fully integrated into the PSSP software and appear similar to regular processor nodes to the PSSP software. This is in contrast to

dependent nodes, such as the SP Switch Router, which are mostly ignored by the PSSP software. Also, unlike the SP Switch Router, the SP-attached server must be assigned a frame number.

The SP-attached server's frame number needs to be manually configured because the server has frame-like characteristics but does not have a frame supervisor card. When assigning the frame number:

- Do not make the SP-attached server the first frame in the SP system.
- The assigned frame number cannot be a number that would come between the number assigned to a switch configured frame and the frame numbers assigned to any non-switched expansion frames attached to the switch configured frame.
- You can select any frame number up to the maximum number of frames allowed provided it does not violate the first two points.
- Specify the hardware protocol (SAMI) for the SP-attached server.

For more information on this topic, see Section 4.1.1, "How the SP system views the SP-attached server" on page 74.

Assigning a switch port number

Regardless of whether your SP system is switch configured or if it is a non-switched SP system, you must assign a switch port number to the SP-attached server.

- For a switch configured SP system, this number can be any valid, unused switch port in the SP system.
See Section 4.3.2, "Connecting to the SP Switch" on page 88 for specific details.
- For a non-switched SP system, an unused node slot is required in the SP frame associated with the SP-attached server.
See "Software configuration in a non-switched system" on page 83 for specific details.

Why you need to assign a switch port

With a standard SP frame, once the frame number is assigned, the PSSP software generates the node numbers. Also, a configuration algorithm determines the switch port assignment for any non-switched expansion frame attached to the switch configured frame.

With an SP-attached server, the PSSP software generates the node numbers but cannot assign a switch port number because the software does not see the server as a standard non-switched expansion frame. Because of that, you must assign the switch port number.

Software configuration in a non-switched system

SP-attached servers can be installed in a non-switched SP system. You install and configure SP-attached servers in these systems just as you would in a switch configured system. The only difference occurs in assigning a switch port number to the server. In a switch configured system, the switch port number you assign must be any valid, unused switch port. In a non-switched system, you must calculate the switch port number.

This calculation is required even though a non-switched system does not have a switch, and the SP-attached server does not require an RS/6000 SP System Attachment adapter. In a non-switched system, the switch port number is assigned to the SP-attached server based on availability of a valid, open node slot in the system.

To determine what switch port you must assign to the SP-attached server, use the following formula:

$$\text{switch_port_number} = ((\text{frame_number_of_associated_SP_frame} - 1) \times 16) + \text{assigned_node_slot_number}$$

In this formula, the associated SP frame contains the valid, open node slot you are assigning to the SP-attached server. That frame's number and the number of the open node slot are used to calculate the switch port number even though the SP-attached server is not directly connected with this SP frame.

4.3 Network interface

There is a maximum number of SP-attached servers that can be connected to an SP system. See "Server attachment limitations" on page 75 for more information.

The SP-attached server requires a minimum of four connections with your SP system in order to establish a functional and safe network. If your SP system is configured with an SP Switch, there will be five required connections:

- Three connections are required with the control workstation:
 1. An Ethernet connection to the SP Ethernet for system administration purposes.
See Section 4.3.1.1, “Attaching the SP Ethernet” on page 86 for cable details
 2. A custom RS-232 cable connecting the SP’s control workstation to the SP-attached server’s SAMI port.
Uses IBM supplied 15 m (49 foot) cable.
 3. A second custom RS-232 cable connecting the SP’s control workstation to the SP-attached server’s S1 serial port.
It uses the IBM supplied 15 m (49 foot) cable.
See Section 4.3.1, “Connecting to the control workstation” on page 86 for details
- The fourth connection is a 10 m frame-to-frame electrical ground cable.
The SP-attached server must be connected to the SP frames with an IBM supplied grounding cable. This cable is supplied with the SP system when you order F/C 9122 and F/C 9123.
The frame-to-frame ground is required in addition to the SP-attached server electrical ground. The frame-to-frame ground maintains the SP and SP-attached server at the same electrical potential.
- The fifth connection is required if the SP system is switch configured.
In these systems, the SP-attached server must also have an optional RS/6000 SP System Attachment adapter (RS/6000 F/C 8396) installed. This adapter uses a 10 m SP Switch cable to connect to a valid switch port on an SP Switch. See Section 4.3.2, “Connecting to the SP Switch” on page 88 for details.

These five network connections are shown in Figure 13.

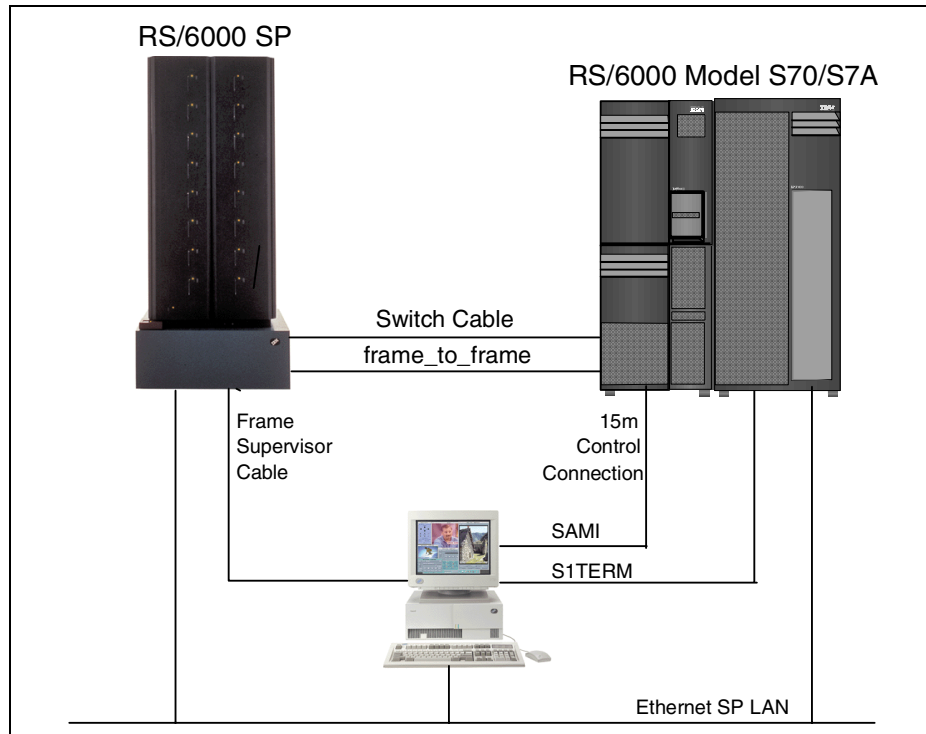


Figure 13. SP-attached server network connections

Cable length limitation

Placement of the SP-attached server is limited by the length of cables supplied with the SP-attached server:

- The 10 m (33 ft.) cable length of both the frame-to-frame ground and the RS/6000 SP System Attachment adapter cable
- The 15 m (49 ft.) cable length of the RS-232 cables
- BNC Ethernet cable is limited to 15 m (49 ft.)

Remember, approximately 3 m (10 feet) of cable are needed for vertical cable runs between equipment. Therefore, the SP-attached server should be no more than 7 m (23 feet) from the SP frame and no more than 12 m (46 feet) from the control workstation.

4.3.1 Connecting to the control workstation

The SP-attached server requires three connections with the control workstation. These connections are:

1. An Ethernet connection to the SP Ethernet for system administration purposes
2. A custom RS-232 cable connecting the SP's control workstation to the SP-attached server's SAMI port
3. A second custom RS-232 cable connecting the SP's control workstation to the SP-attached server's S1 serial port

4.3.1.1 Attaching the SP Ethernet

Only two Ethernet adapters are supported for SP Ethernet communication. These adapters are:

1. F/C 2985 10 MB BNC/RJ-45 Ethernet (ordered with the 7017 Enterprise Server)
 - Requires one F/C 9222 for each SP-attached server (ordered on the SP, configures BNC connection to server).
When F/C 9222 is ordered, a 15 m BNC Ethernet cable is included for the SP.
 - For more information on this adapter, see Section 9.1.13, "Ethernet 10 MB BNC (F/C 2985)" on page 157.
2. F/C 2987 10 MB AUI/RJ-45 Ethernet (ordered with the 7017 Enterprise Server)
 - Requires one F/C 9223 for each SP-attached server (ordered on the SP, configures twisted pair connection to server).
When F/C 9223 is ordered, the customer must supply all twisted pair Ethernet cables.
 - For more information on this adapter, see Section 9.1.14, "Ethernet 10 MB AUI (F/C 2987)" on page 158.

SP Ethernet connection

- When choosing between F/C 2985 and F/C 2987, the adapter that you choose must match the configuration of the SP Ethernet on your SP system. These adapters must be placed in the en0 position of the SP-attached server.
- The en0 position is the lowest numbered Ethernet bus slot in the first I/O tower.
- Note that although the 10/100 MB Ethernet adapter (F/C 2968) is an SP supported Ethernet adapter, it is not supported for SP Ethernet communication. F/C 2968 may be used in other slots in the SP-attached server, but it cannot be used in the en0 position

Ethernet adapter restrictions

If you are attaching an existing 7017 Enterprise Server to an SP System, you must place an SP Ethernet adapter in the en0 position inside the SP-attached server. Because the Ethernet adapter in this slot must be configured for SP communications, any non-SP supported Ethernet adapter that has been placed into this slot must be removed.

Also, even if the Ethernet adapter in en0 is either F/C 2985 or F/C 2987, the adapter must be deconfigured and reconfigured as an SP Ethernet adapter.

4.3.1.2 Attaching the RS-232

Two RS-232 connections must be made from the control workstation to the SP-attached server. These attachments go to the SP-attached server's:

1. SAMI port
 - Located in the control panel on the front of the CEC
 - Uses an IBM supplied 15 m custom RS-232 cable
2. S1 serial port
 - Located on the rear of the primary (first) I/O tower
 - Uses an IBM supplied 15 m custom RS-232 cable

Since the SP-attached server requires multiple RS-232 connections, you will need to use a multiport asynchronous adapter inside the control workstation. See Section 8.2.3.1, "Serial port adapters" on page 128, for a listing of the available adapters, RANs, and cables.

Note that the 16-port asynchronous adapter (F/C 2955) used in MCA control workstations is not compatible with the SP-attached server.

4.3.2 Connecting to the SP Switch

If your SP system is configured with an SP Switch, the SP-attached server requires a connection between the SP Switch and the server. To make this connection, your system will require the RS/6000 SP System Attachment adapter (RS/6000 F/C 8396). This adapter occupies three media card slots in the SP-attached server. See Section 4.2.2.1, “Placement restrictions” on page 78 for specific details on configuring the RS/6000 SP System Attachment adapter.

Once the adapter has been installed in the SP-attached server, the 10 m switch cable F/C 9310 must be attached to a valid switch port on the SP Switch. The general steps for choosing a valid SP Switch port are outlined here.

4.3.2.1 Selecting a valid switch port

In a switch configured SP system, each SP-attached server requires one RS/6000 SP System Attachment adapter. Only one switch adapter may be used per server. Each of these RS/6000 SP System Attachment adapters will require a valid unused switch port in the SP system. A valid unused switch port is a switch port that meets the rules for configuring frames and switches.

There are two basic sets of rules for choosing a valid switch port:

1. Rules for selecting a valid switch port associated with an empty node slot.
2. Rules for selecting a valid switch port associated with an unused node slot created by a wide or high node. These node slots are either the second half of a wide node or one of the last three positions of a high node.

Examples of using an empty node slot position

One example of using an empty node slot position is a single frame system with a switch and 14 thin nodes located in slots 1 through 14. This system has two unused node slots in position 15 and 16. These two empty node slots have corresponding switch ports that provide valid connections for the RS/6000 SP System Attachment adapter.

Another example is a two-frame system with one switch. The first frame is fully populated with eight wide nodes. The second frame has three wide nodes in system node positions 17, 19, and 21. The only valid switch ports in this configuration would be those switch ports associated with system node numbers 23, 25, 27, 29, and 31 in the second frame.

In a four-frame system with one switch and fourteen high nodes, there will only be two empty node positions. In this example, the first three frames are fully populated with four high nodes in each frame. The last frame has two

high nodes and two empty high node slots. This means the system has two valid switch ports associated with system node numbers 57 and 61.

Examples of using node slot positions within a wide or high node

The first example is a single frame with a switch and eight wide nodes. These wide nodes occupy the odd numbered node slots. Therefore, all of the even number slots are said to be unoccupied and would have valid switch ports associated with them. These ports may be used for an RS/6000 SP System Attachment adapter.

A second example is a single frame system with a switch, twelve thin nodes in slots 1 through 12, and a high node in slot 13. A high node occupies four slots but uses only one switch port. Therefore, the only valid switch ports in this configuration are created by the three unused node slots occupied by the high node. In other words, the switch ports are associated with node slots 14, 15, and 16.

Chapter 5. Clustered Enterprise Server System

The Clustered Enterprise Server (CES) system is a group of one to sixteen IBM RS/6000 7017 Enterprise Servers and an RS/6000 control workstation (CWS), all running PSSP software. Clustered Enterprise Servers benefit from the PSSP software single point of service and provide the scalability needed for e-business applications.

5.0.1 Clustered Enterprise Server overview

Enterprise Servers 7017-S70, S7A, and S80 use *RS/6000* feature codes, *not RS/6000 SP* feature codes. The RS/6000 feature codes associated with the Clustered Enterprise Server refer to the cable connections that attach the servers to the control workstation.

5.0.2 Clustered Enterprise Server installation requirements

There are several requirements for hardware and software that must be met before you can place the Clustered Enterprise Servers into service as follows:

- System requirements
- Physical requirements
- Network media card requirements
- Software requirements
- Network cables

5.0.3 Clustered Enterprise Server system requirements

The following requirements must be met before you can place Clustered Enterprise Servers into service:

1. Your CES system must be operating with at least PSSP 3.2 and AIX 4.3.3 software.
2. You will need three control workstation cable connections. Some of the cables supplied with the CES have limited length. Those lengths, along with the vertical parts of the cable runs, limit the location of the servers in relation to the other system equipment.
3. If you plan to attach an **existing** Enterprise Server, and that server is connected to an **IPV6 network**, you must remove the server from that network before including it in the Clustered Enterprise Server system.

PCI adapters in an **existing** Enterprise Server must be SP-supported.

5.0.4 Clustered Enterprise Server network media card requirements

Each network media card requires one media card slot in the Enterprise Server. All network adapters in the Clustered Enterprise Server use PCI architecture.

PCI adapter restriction

Only SP system supported PCI adapters can be used in a 7017 Enterprise Server when it is used as a Clustered Enterprise Server. Thus, if you attach an existing RS/6000 7017 Enterprise Server to a Clustered Enterprise Server system, you must remove any non-SP system supported PCI adapters.

Note

Rules for supported PCI adapters, such as the following, can be found in the RS/6000 7017 Enterprise Server documentation:

- Required adapters (including minimum requirements)
- Maximum quantity of each adapter allowed
- Bus placement restrictions

If an adapter does not appear in the list, it is not supported for the Clustered Enterprise Server. If you plan to use an existing Enterprise Server and any of its installed adapters do not appear in the list, they must be removed before the server can be attached to the CES system.

Note

F/C 2985 and 2987 have placement restrictions when used as the SP-LAN adapter.

5.0.5 Clustered Enterprise Server software requirements

The Clustered Enterprise Servers and the control workstation require the following software levels:

- PSSP 3.2 (or later)
- AIX 4.3.3 (or later)

Each Clustered Enterprise Server requires its own PSSP license. PSSP software is available in the following formats:

- F/C 5800 (4 mm tape)

- F/C 5801 (8 mm tape)
- F/C 5802 (DC-ROM)

For details on software requirements of the Clustered Enterprise Server, see RS/6000 SP: Planning Volume 2, Control Workstation and Software Environment.

5.0.6 Planning the Clustered Enterprise Server network

Each Clustered Enterprise Server requires a minimum of three cable connections with the control workstation to establish a functional and safe network as follows:

1. An Ethernet cable connecting to the SP-LAN for system administration purposes
2. Two custom RS-232 cables connecting the CWS to both the server SAMI port and to the S1 serial port

Clustered Enterprise Server placement limitations

The location of the Clustered Enterprise Servers is limited by the length of the IBM supplied 15 m (49 ft.) RS-232 and BNC Ethernet cables. Approximately 3 m (10 ft.) of cable is needed for the vertical portions of these cable runs. Thus, the servers can be no more than 12 m (40 ft.) from the control workstation.

5.0.6.1 Attaching the SP-LAN Ethernet hardware

Three Ethernet adapters ordered with the Enterprise Servers are supported for SP-LAN Ethernet communication. These adapters are:

- **Twisted-pair** cable connection
 - 10/100 Ethernet 10BaseTX adapter (F/C 2968)
 - 10 MB AUI/RJ-45 Ethernet adapter (F/C 2987)

Note that the customer must supply all twisted pair Ethernet cables.
- **BNC** cable connection - 10 MB BNC/RJ-45 Ethernet adapter (F/C 2985)

SP-LAN Ethernet requirements

The adapter you select must match the cable connection configuration of the SP-LAN for your CES system. These adapters must be placed in the en0 position of the Clustered Enterprise Server (the lowest-numbered Ethernet bus slot in the first I/O tower).

Ethernet adapter restrictions

If you plan to attach an existing 7017 Enterprise Server to your system, you must place an SP-LAN Ethernet adapter in the slot en0 position inside the server. Because the Ethernet adapter in this slot must be configured for PSSP communications, any non-supported Ethernet adapter that is in the en0 slot must be removed.

Additionally, if the Ethernet adapter in slot en0 is either of F/C 2968, 2985, or 2987, the adapter must be de-configured and then reconfigured as an SP-LAN Ethernet adapter.

5.0.6.2 Attaching the RS-232 cables

Since the Clustered Enterprise Servers require multiple RS-232 connections, you must use a multiport, asynchronous adapter in the control workstation.

Two RS-232 connections must be made from the control workstation to each Clustered Enterprise Server. These connections go to the following ports on the servers:

1. SAMI port in the control panel on the front of the CEC with an IBM-supplied 15 m (49 ft.) custom RS-232 cable (F/C 3151)
2. S1 serial port on the rear of the primary (first) I/O tower with an IBM-supplied 15 m (49 ft.) custom RS-232 cable (F/C 3150)

Note

The 16-port asynchronous adapter (F/C 2955) used in MCA-type control workstations is not compatible with the Clustered Enterprise Server system.

5.0.6.3 Configuring Service Director

Service Director is a set of IBM software applications supplied with the 7017 Enterprise Servers. Service Director monitors the health of the system.

In a typical Enterprise Server installation, Service Director transmits reports through a modem supplied with the unit. However, when the 7017 Enterprise Server is used as a Clustered Enterprise Server, the modem supplied with the 7017 is not used. In this installation, the CES acts like a system node and forwards its Service Director messages to the system. When the system receives messages from the Clustered Enterprise Server, the messages are transmitted through the Service Director modem.

To configure Service Director for the Clustered Enterprise Server you must perform the following:

1. Configure the Clustered Enterprise Server as a Machine Type 7017 in Service Director. You must do this manually.
2. Configure Service Director on each Clustered Enterprise Server to forward messages to the system. The modem supplied with the 7017 Enterprise Server is not used.
3. Configure Service Director on the system to forward messages received from the Clustered Enterprise Server. The Service Directory modem for the CES system is attached to the control workstation.

Chapter 6. SP Switch Routers

The IBM 9077 SP Switch Router is a licensed version of the Ascend GRF switched IP router that has been enhanced for direct connection to the SP Switch. IBM remarkets models of the GRF that connect to the SP Switch as the SP Switch Router:

- IBM 9077 SP Switch Router model 04S (9077-04S) is based on Ascend GRF 400.
- IBM 9077 SP Switch Router model 16S (9077-16S) is based on Ascend GRF 1600.

To connect SP Switch Router to an SP system, the following adapter must be installed:

- SP Switch Router adapter (F/C 4021)

Figure 14 shows IBM 9077 SP Switch Routers.



Figure 14. IBM 9077 SP Switch Router model 04S (left) and 16S (right)

6.1 Overview

A physical dependent node, such as an RS/6000 SP Switch Router (Machine Type 9077), may have multiple logical dependent nodes, one for each dependent node adapter it contains. If a dependent node, such as the SP Switch Router, contains more than one dependent node adapter, it can route data between SP systems or system partitions. For the RS/6000 SP Switch Router, this card is called a Switch Router adapter (F/C 4021). Data

transmission is accomplished by linking the dependent node adapters in the switch router with the logical dependent nodes located in different SP systems or system partitions.

In addition to the four major dependent node components (see Section 1.2.3.1, “Dependent nodes” on page 13), the SP Switch Router (dependent node) has a fifth optional category of components. These components are networking cards that fit into slots in the SP Switch Router. In the same way that the SP Switch Router adapter connects the SP Switch Router directly to the SP Switch, these networking cards enable the SP Switch Router to be directly connected to an external network. The following networks can be connected to the RS/6000 SP Switch Router using available media cards:

- Ethernet 10/100 Base-T
- FDDI
- ATM OC-3c (single or multimode fiber)
- SONET OC-3c (single or multimode fiber)
- ATM OC-12c (single or multimode fiber)
- HiPPI
- HSSI

You can find a full list of these networking cards in Section 6.2.3, “Network media card requirements” on page 103.

Figure 15 on page 99 shows the SP Switch Router configuration example. The SP Switch Router can be used for high-speed network connections or system scaling using HIPPI backbones or other communications subsystems, such as ATM or 10/100 Ethernet.

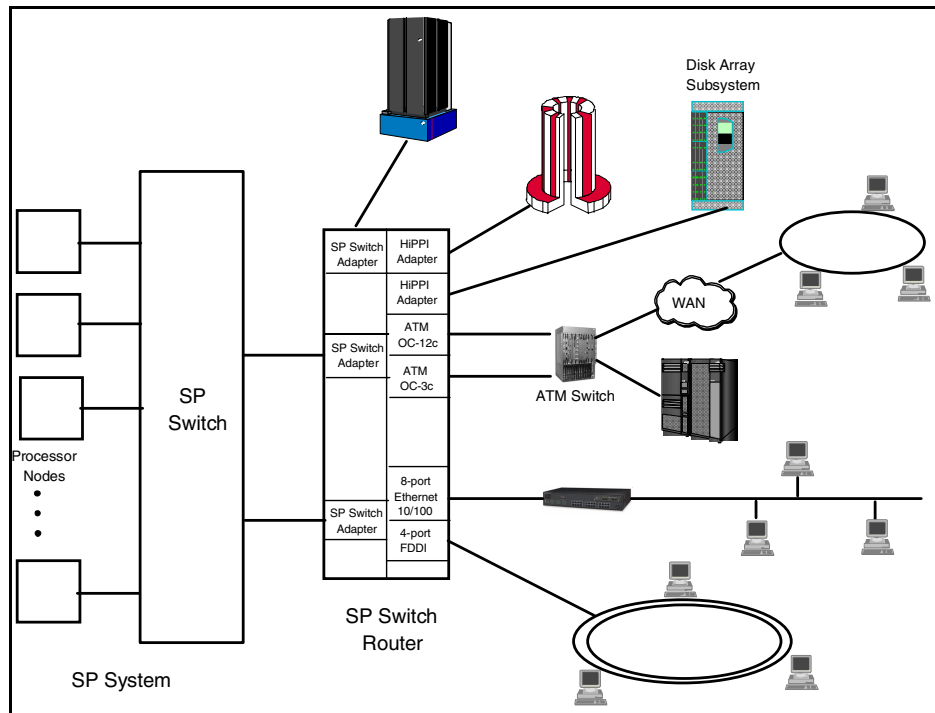


Figure 15. SP Switch Router configuration

Although you can equip an SP node with a variety of network adapters and use the node to make your network connections, the SP Switch Router with the Switch Router adapter and optional network media cards offers many advantages when connecting the SP to external networks.

- Each media card contains its own IP routing engine with separate memory containing a full route table of up to 150,000 routes. Direct access provides much faster lookup times compared to software driven lookups.
- Media cards route IP packets independently at rates of 60,000 to 130,000 IP packets per second. With independent routing available from each media card, the SP Switch Router gives your SP system excellent scalability characteristics.
- The SP Switch Router has dynamic network configuration to bypass failed network paths using standard IP protocols.
- Using multiple Switch Router adapters in the same SP Switch Router, you can provide high performance connections between system partitions in a single SP system or between multiple SP systems.

- A single SP system can also have more than one SP Switch Router attached to it, thus, further insuring network availability.
- Media cards are hot swappable for uninterrupted SP Switch Router operations.
- Each SP Switch Router has redundant (N+1) hot swappable power supplies.

Two versions of the RS/6000 SP Switch Router can be used with the SP Switch:

- The Model 04S offers four media card slots.
- The Model 16S offers sixteen media card slots.

Except for the additional traffic capacity of the Model 16S, both units offer similar performance and network availability.

Figure 16 illustrates, from a rear view, SP Switch Router model 04S with four optional network media cards.

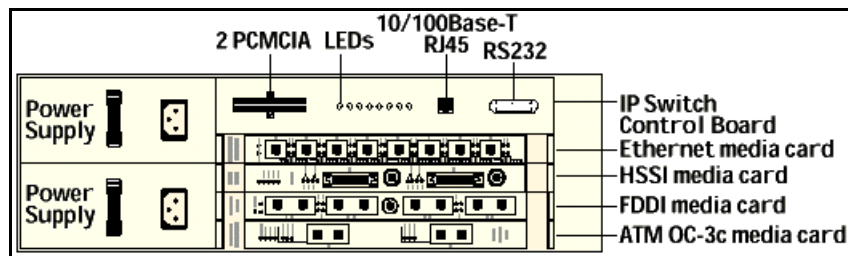


Figure 16. Rear view of SP Switch Router model 04S

Figure 17 illustrates, from a rear view, SP Switch Router model 16S with nine optional network media cards.

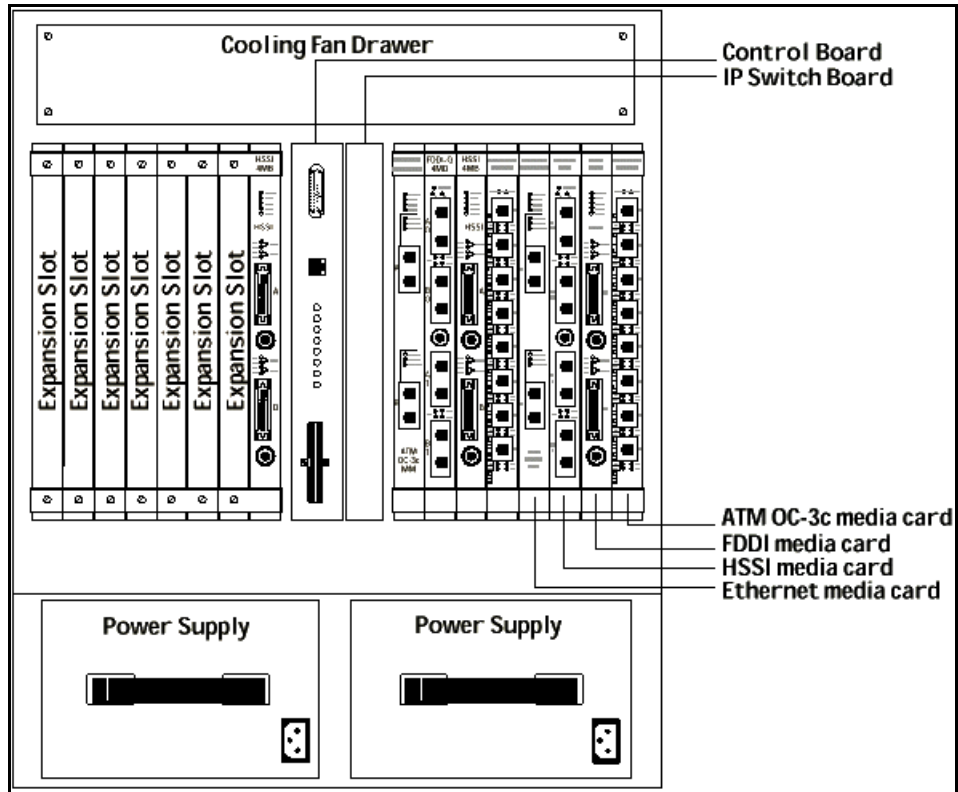


Figure 17. Rear view of SP Switch Router model 16S

6.2 Installation requirements

There are several requirements for hardware and software that must be met before you can place the RS/6000 SP Switch Router into service with your SP system. These requirements are in the following categories:

- System requirements
- Switch adapter requirements
- Network media card requirements
- Software requirements

6.2.1 System requirements

In addition to the SP Switch Router, the following requirements must be met before you can place the router into service:

- You must have at least one SP Switch Router adapter.
- You will need a VT100 compatible terminal with an RS-232 cable and null modem for initial configuration of the SP Switch Router.
- You will need a 10Base-T connection between your SP control workstation and the SP Switch Router. If your control workstation uses 10Base-2 Ethernet, you must also supply a 10Base-T to 10Base-2 bridge.
- Your SP system must be equipped with either an SP Switch (F/C 4011) or an SP Switch-8 (F/C 4008).
- The SP Switch Router includes 128 MB of memory. This memory is used for storing routing information for 199,485 static and dynamic routes. F/C 1114 can be used to increase memory capacity on 64 MB increments up to a maximum of 256 MB, which allows up to 150,000 static route entries and 521,730 dynamic route prefixes per memory card.
- You must attach a frame-to-frame ground between the SP system and the SP Switch Router using the IBM supplied cable in order to maintain both systems at the same electrical potential.

6.2.2 SP Switch Router adapter (F/C 4021)

If you are placing an SP Switch Router into an SP system, you must install an SP Switch Router adapter (F/C 4021) in the SP Switch Router. The SP Switch Router adapter requires the following:

- One valid, unused switch port on the SP Switch corresponding to a legitimate node slot in your SP configuration.

A legitimate node slot may be empty, the second half of a wide node, or one of the last three positions of a high node, provided that node slot satisfies the other rules for configuring nodes in an SP system. For example, if you have a frame with 16 thin nodes installed, you cannot attach a Switch Router adapter to that frame until you remove a node and delete its configuration from the system image.
- One media card slot in the SP Switch Router. The RS/6000 SP Switch Router Model 04S has the capacity for a total of four SP Switch Router adapters and network media cards in any combination suiting the needs of your SP system. The RS/6000 SP Switch Router Model 16S has the capacity for a total of 16 SP Switch Router adapters and network media cards in any combination.

6.2.3 Network media card requirements

Each network media card requires one media card slot in the RS/6000 SP Switch Router. Remember, the network media cards use the same slots as the SP Switch Router adapters.

The following network media cards shown in Figure 14 are available as options for the SP Switch Router.

Table 14. SP Switch Router network media cards and other options

F/C	Description
1101	ATM OC3, two port SM fiber
1102	ATM OC3, two port MM fiber
1103	SONET/IP OC3, one port MM fiber
1104	SONET/IP OC3, one port SM fiber
1105	ATM OC12, one port SM fiber
1106	FDDI, four port MM fiber
1107	Ethernet 10/100Base-T, eight port
1108	HIPPI, one port
1109	HSSI, two port
1112	Ethernet 10/100Base-T, four port
1113	Blank faceplate
1114	64 MB DRAM SIMM
1115	ATM OC12, one port MM fiber
4021	SP Switch Router adapter ¹
9310	SP Switch Router adapter cable, 10 meter option (includes 10 m frame-to-frame ground cable)
9320	SP Switch Router adapter cable, 20 meter option (includes 20 m frame-to-frame ground cable)
¹ Choice of either F/C 9310 or F/C 9320 is included with each F/C 4021.	

For a brief description of these features, see Section 9.3, “SP Switch Routers network media cards” on page 176.

6.2.4 Software requirements

The SP Switch Router requires an SP system operating with PSSP 2.3 (or later) with the appropriate APAR level and AIX 4.2.1 (or later) on the primary and backup nodes for the SP Switch and on the control workstation.

If the SP Switch Router is used in an SP partition where there are nodes operating at lower than the required level of PSSP and AIX, you will have to apply service updates to the software operating on those nodes.

6.3 Network interface

The RS/6000 SP Switch Router (Machine Type 9077) requires a minimum of three connections with your SP system in order to establish a functional and safe network. These connections are:

1. A network connection with the control workstation.

The SP Switch Router must be connected to the control workstation for system administration purposes. This connection may be either:

- A direct Ethernet connection between the SP Switch Router and the control workstation
- An Ethernet connection from the SP Switch Router to an external network, which then connects to the control workstation

See Section 6.3.1, “Connecting to the control workstation” on page 105 for more information.

2. A connection between an SP Switch Router adapter and the SP Switch.

The SP Switch Router transfers information into and out of the processor nodes of your SP system. The link between the SP Switch Router and the SP processor nodes is implemented by an SP Switch Router adapter (F/C 4021) and a switch cable connecting the Switch Router adapter to a valid switch port on the SP Switch. See Section 6.3.2, “Connecting to the SP Switch” on page 107 for more information.

3. A frame-to-frame electrical ground.

The SP Switch Router frame must be connected to the SP frame with a grounding cable. This frame-to-frame ground is required in addition to the SP Switch Router electrical ground. The purpose of the frame-to-frame ground is to maintain the SP and SP Switch Router systems at the same electrical potential.

Both the SP Switch cable and the grounding cable are shipped with each SP Switch Router adapter. The recommended cable for connecting the SP

Switch Router adapter to the SP Switch is 10 meters long (F/C 9310). An optional 20 meter cable (F/C 9320) is also available for the SP Switch connection. A frame-to-frame ground cable the same length as the SP Switch cable is included with both F/C 9310 and F/C 9320.

Network interface of the SP Switch Router is shown in Figure 18. VT100 terminal is optional and required for initial configuration time.

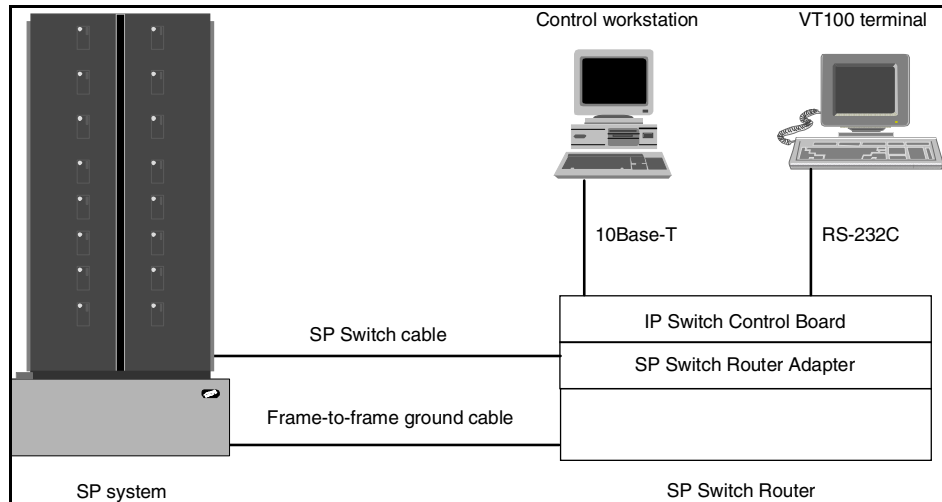


Figure 18. SP Switch Router network interface

The following sections describe how to connect the SP Switch Router to the control workstation and how to connect the SP Switch Router adapter to a valid SP Switch port.

6.3.1 Connecting to the control workstation

Although a dependent node, such as the SP Switch Router, does not function like a processor node, it must be administered by the SP system as if it were a processor node. Therefore, the SP Switch Router must be attached to the control workstation. All SP administrative connections are made to the SP Switch Router using the port on the Router's control board. From the SP Switch Router's control board, the connection to the control workstation is made using one of the following methods:

1. If the control workstation is connected to the SP system through a 10Base-2 (thin coax) network, the SP Switch Router may be connected to the network through a customer supplied 10Base-T to 10Base-2 hub (or bridge), such as the IBM 8222 Ethernet Workgroup Hub.

All coax and twisted pair Ethernet cables must be supplied by the customer.

2. If the control workstation is connected to the SP system through a twisted pair (TP) Ethernet LAN, the SP Switch Router may be connected to an available port on the Ethernet hub (switch).

All coax and twisted pair Ethernet cables must be supplied by the customer.

3. The RS/6000 SP Switch Router may also be connected to an additional 10Base-T adapter (such as F/C 2992) that has been installed directly in the control workstation for this purpose. If you decide to use this method, you must set up a separate Ethernet subnet for the SP Switch Router.

When using separate 10Base-T adapters for the control workstation connection, in addition to the 10Base-T adapter, you must also supply a twisted pair Ethernet cable with a crossed connection appropriate for use between two network interface cards.

4. The SP Switch Router may also be indirectly connected to the control workstation using an external network. In this configuration, the Ethernet connection from the router's control board is attached to external network equipment. The external network connection to the control workstation may be:

- A separate (non-SP Ethernet) Ethernet
- ATM
- FDDI

6.3.1.1 Connecting to multiple SP Systems

If you are planning to connect one SP Switch Router to multiple, independent SP Systems, you will need:

- One SP Switch Router adapter for each SP system being connected.
- A Switch Router adapter cable to connect each of the adapters to an SP Switch located in each of the SP systems.
- An Ethernet connection from the SP Switch Router's control board to an external network.

The Router's control board Ethernet connection is de0 and uses 10/100BaseT Ethernet.

- Connections from the external network must attach to the control workstations administering each SP system. The external networks may be:

- Other Ethernets
- ATM
- FDDI
- Frame to frame grounds are required.

Valid control workstation connections

Other methods can be used to make the connection between the Router control board and all control workstations used in the SP systems. Any method providing the ability to ping SP control workstations from the Router control board will provide a valid path.

6.3.2 Connecting to the SP Switch

In addition to the control workstation Ethernet connection, the RS/6000 SP Switch Router requires a connection between the SP Switch and the SP Switch Router. To make this connection, your system will require the SP Switch Router adapter (F/C 4021). This adapter occupies one media card slot in the attached SP Switch Router.

6.3.2.1 Selecting a valid switch port

An SP Switch Router adapter in the SP Switch Router may be attached to an SP Switch to improve throughput of data coming into and going out of the RS/6000 SP system. Each SP Switch Router adapter in the RS/6000 SP Switch Router will require a valid unused switch port in the SP system. A valid unused switch port is a switch port that meets the rules for configuring frames and switches.

There are two basic sets of rules for choosing a valid switch port:

1. Rules for selecting a valid switch port associated with an empty node slot.
2. Rules for selecting a valid switch port associated with an unused node slot created by a wide or high node. These node slots are either the second half of a wide node or one of the last three positions of a high node.

Examples of using an empty node slot position

One example of using an empty node slot position is a single frame system with an SP Switch and 14 thin nodes located in slots 1 through 14. This system has two unused node slots in position 15 and 16. These two empty node slots have corresponding switch ports that provide valid connections for the SP Switch Router adapter.

Another example is a two-frame system with one switch. The first frame is fully populated with eight wide nodes. The second frame has three wide nodes in system node positions 17, 19, and 21. The only valid switch ports in this configuration would be those switch ports associated with system node numbers 23, 25, 27, 29, and 31 in the second frame.

In a four-frame system with an SP Switch and fourteen high nodes, there will be only two empty node positions. In this example, the first three frames are fully populated with four high nodes in each frame. The last frame has two high nodes and two empty high node slots. This means the system has two valid switch ports associated with system node numbers 57 and 61.

Examples of using node slot positions within a wide or high node

The first example is a single frame with an SP Switch and eight wide nodes. These wide nodes occupy the odd numbered node slots. Therefore, all of the even number slots are said to be unoccupied and would have valid switch ports associated with them. These ports may be used for an SP Switch Router adapter.

A second example is a single frame system with an SP Switch, twelve thin nodes in slots 1 through 12, and a high node in slot 13. A high node occupies four slots but only uses one switch port. Therefore, the only valid switch ports in this configuration are created by the three unused node slots occupied by the high node. In other words, the switch ports are associated with node slots 14, 15, and 16.

Chapter 7. SP Switch network

The SP Switch is the foundation of efficient message passing between SP nodes. Its characteristics, and the communication subsystem that supports it, differentiate the SP system from clusters of workstations as well as competitive parallel systems. The SP Switch provides the message passing capability that allows all processor nodes in the system to send messages simultaneously. The SP Switch chooses the indirect network approach based on the following:

- In an indirect network, including the SP Switch, bisectional bandwidth scales linearly with the number of processor nodes in the system.
- Most indirect networks, including the SP Switch, can support an arbitrarily large interconnection network while maintaining a fixed number of ports per switch.
- Deadlock will not occur in most indirect networks as long as the packet travels along any shortest-path route. In the SP Switch, there are typically at least four shortest-path routes between any two processor nodes.
- Indirect networks, including the SP Switch, allow packets that are associated with different messages to be spread across multiple paths, thus reducing the occurrence of hot spots.

The hardware component that supports this communication network consists of two basic components: The SP Switch adapter and the SP Switch board. There is one SP Switch adapter per processor node and generally one SP Switch board per frame. This setup provides connections to other processor nodes. Also, the SP system allows switch boards-only frames that provide switch-to-switch connections and greatly increase scalability.

Figure 19 shows the front view of the SP Switch board:

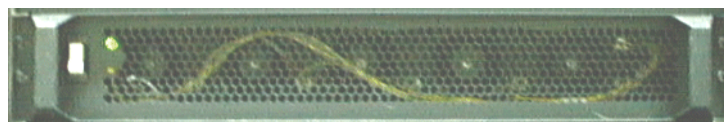


Figure 19. Front view of SP Switch board

7.1 Overview

This section discusses the hardware design that makes up the SP Switch network: The SP Switch link, the SP Switch port, the SP Switch chip, the SP

Switch adapter, and the SP Switch board. The SP Switch link itself is the physical cable connecting two SP Switch ports. The SP Switch ports are hardware subcomponents that can reside on an SP Switch adapter that is installed in a node or on an SP Switch chip that is part of an SP Switch board.

7.1.1 SP Switch board

An SP Switch board contains eight SP Switch chips that provide connection points for each of the nodes to the SP Switch network as well as for each of the SP Switch boards to the other SP Switch boards. The SP Switch chips each have a total of eight SP Switch ports, which are used for data transmission. The SP Switch ports are connected to other switch ports via a physical SP Switch link.

In summary, there are 32 external SP Switch ports in total. Of these, 16 are available for connection to nodes, and the other 16 to other SP Switch boards. The SP Switch board is mounted in the base of the SP Frame above the power supplies.

A schematic diagram of the SP Switch board is shown on Figure 20.

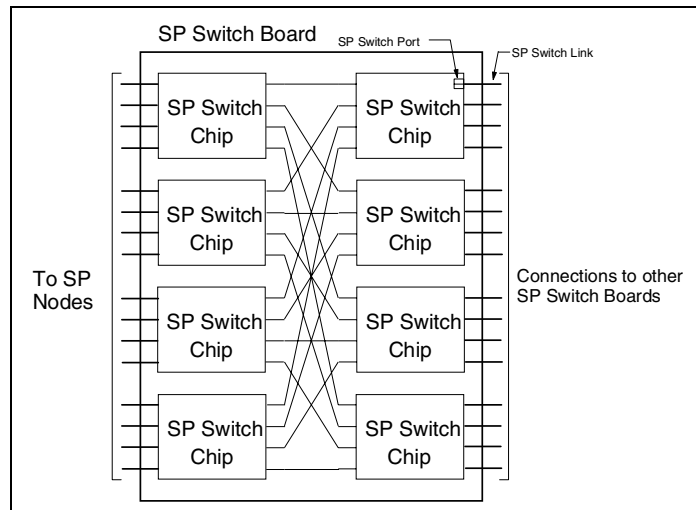


Figure 20. SP Switch board

SP Switch link

An SP Switch link connects two switch network devices. It contains two channels carrying packets in opposite directions. Each channel includes:

- Data (8 bits)

- Data validation (1 bit)
- Token signal (1 bit)

The first two elements here are driven by the transmitting element of the link, while the last element is driven by the receiving element of the link.

SP Switch port

An SP Switch port is part of a network device (either the SP Switch adapter or SP Switch chip) and is connected to other SP Switch ports through the SP Switch link. The SP Switch port includes two ports (input and output) for full duplex communication.

The relationship between the SP Switch chip link and the SP Switch chip port is shown in Figure 21.

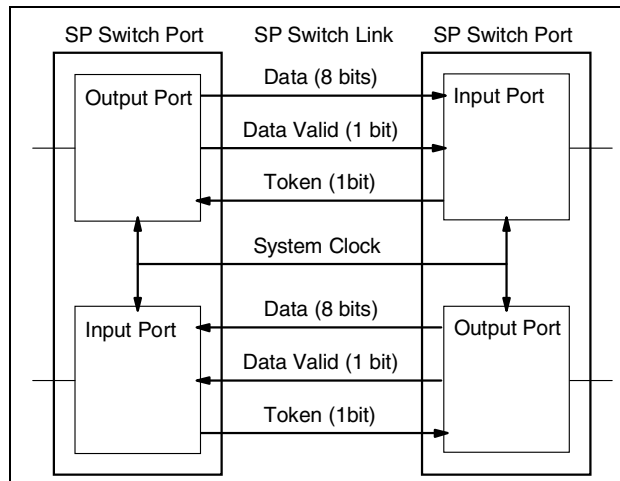


Figure 21. SP Switch chip link and SP Switch chip port

SP Switch chip

An SP Switch chip contains eight SP Switch ports, a central queue, and an unbuffered crossbar, which allows packets to pass directly from receiving ports to transmitting ports. These crossbar paths allow packets to pass through the SP Switch (directly from the receivers to the transmitters) with low latency whenever there is no contention for the output port. As soon as a receiver decodes the routing information carried by an incoming packet, it asserts a crossbar request to the appropriate transmitter. If the crossbar request is not granted, it is dropped (and, hence, the packet will go to the central queue). Each transmitter arbitrates crossbar requests on a least recently served basis. A transmitter will honor no crossbar request if it is

already transmitting a packet or if it has packet chunks stored in the central queue. Minimum latency is achieved for packets that use the crossbar.

A schematic diagram of the SP Switch chip is shown in Figure 22.

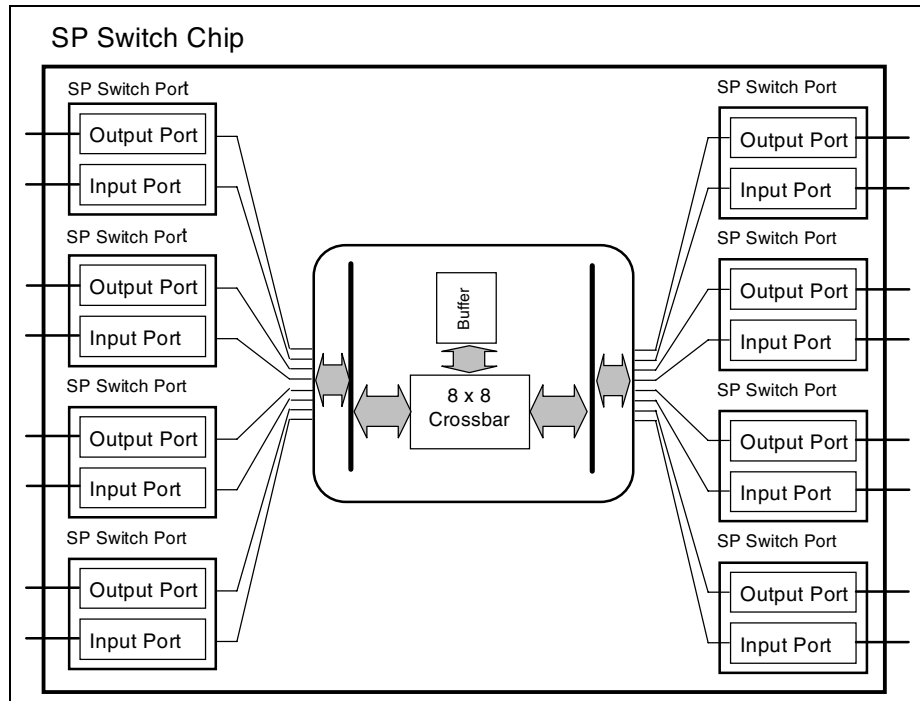


Figure 22. SP Switch chip

7.1.2 SP Switch adapter

Another network device that uses an SP Switch port is the SP Switch adapter. An SP Switch adapter includes one SP Switch port that is connected to an SP Switch board. The SP Switch adapter is installed in an SP node.

External nodes such as the 7017-S70, 7017-S7A and 7017-S80, are based on standard PCI bus architecture. If these nodes are to be included as part of an SP switch network, then the switch adapter needs to be installed in these nodes.

Figure 23 shows a schematic diagram of the SP Switch adapter.

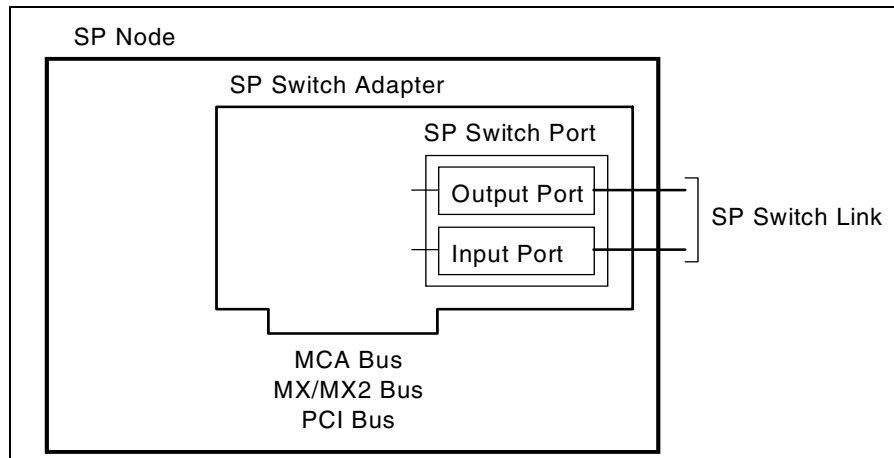


Figure 23. SP Switch adapter

The PCI nodes have a unique internal bus architecture that allows the SP Switch adapters installed in these nodes to see increased performance compared with previous node types.

A conceptual diagram illustrating this internal bus architecture is shown in Figure 24.

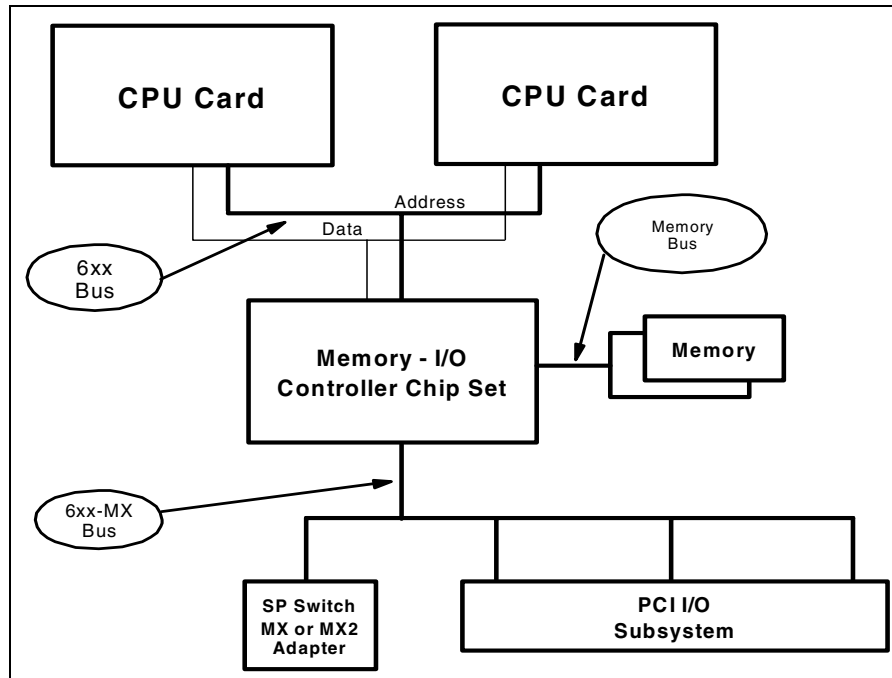


Figure 24. Internal bus architecture for PCI nodes

These nodes implement the PowerPC MP System Bus (6xx bus). In addition, the memory-I/O controller chip set includes an independent, separately clocked mezzanine bus (6xx-MX) to which three PCI bridge chips and the SP Switch MX or MX2 adapter are attached. The major difference between these node types is the clocking rates for the internal buses. The SP Switch adapters in these nodes plug directly into the MX bus. They do not use a PCI slot. The PCI slots in these nodes are clocked at 33 MHz. In contrast, the MX bus is clocked at 50 MHz in the 332 MHz SMP nodes and at 60 MHz in the POWER3 SMP nodes. Thus, substantial improvements in the performance of applications using the switch can be achieved.

7.1.3 SP Switch network

The SP Switch network in a single frame of an SP is illustrated in Figure 25 on page 115. In one SP frame, there are 16 nodes (maximum) equipped with SP Switch adapters and one SP Switch board. Sixteen node SP Switch adapters are connected to 16 of 32 SP Switch ports in the SP Switch board.

The remaining 16 SP Switch ports are available for connection to other SP Switch boards.

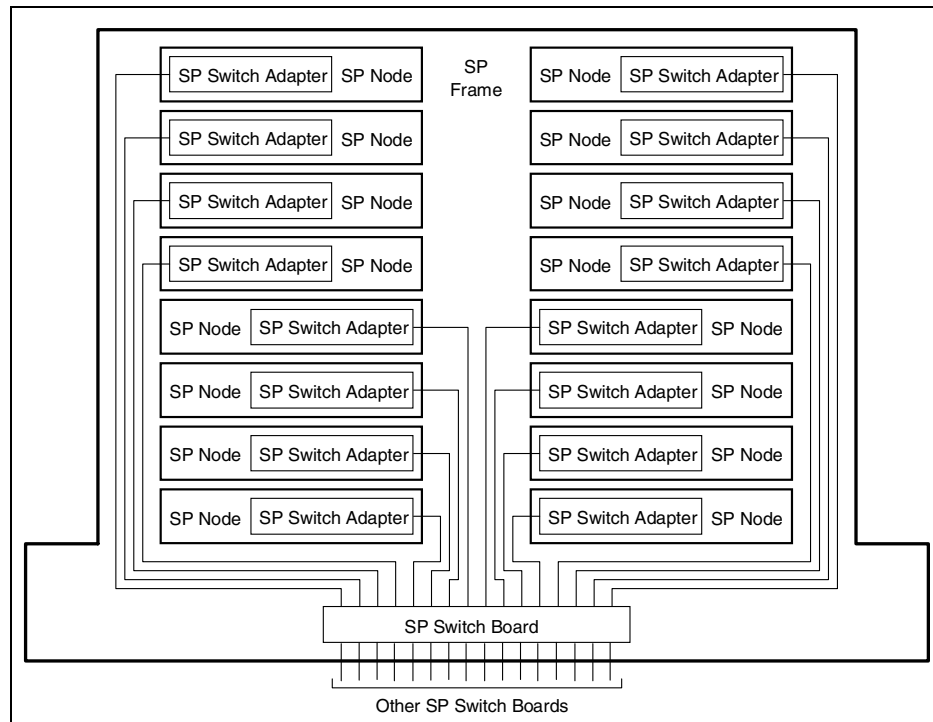


Figure 25. SP Switch network

7.2 SP Switch products

This section describes the SP Switch products, which include:

- SP Switch2 (F/C 4012)
- SP Switch (F/C 4011)
- SP Switch-8 (F/C 4008)
- SP Switch2 Adapter (F/C 4025)
- SP Switch adapter (F/C 4020, 4022, and 4023)
- Special SP Switch adapter (F/C 4021 and RS/6000 F/C 8396)
- SP Switch2 Interposer card (F/C 4032)

7.2.1 SP Switch2 (F/C 4012)

Like the SP Switch, the SP Switch2 switch has 32 ports; 16 ports for switch-to-node connections and 16 ports for switch-to-switch connections. Management of the SP Switch2 network is the same as SP Switch, using service packets over the switch plane. This section describes some of the most important hardware characteristics of the SP Switch2 hardware.

TOD synchronization

The SP Switch2 is designed with multiple clock sources in the system. There is no master clock unlike HPS or SP Switch. Therefore, instead of having clock selection logic, SP Switch2 switch has a separate oscillator for each switch chip. The TOD (Time Of Day) logic on the SP Switch2 has also been significantly redesigned to simplify both coding and system requirements, as well as to improve the accuracy of the TOD across the system.

J-TAG interface

An additional interface, J-TAG interface, is added to the SP Switch2 switch. This new interface will allow the supervisor to perform new functions such as writing initialization data to the switch chips, and reading error status information back from the switch chip.

Adaptive routing

To support the double-port, a number of routing enhancements were added in SP Switch2, namely *adaptive routing* and *multicast packets*. Adaptive routing will allow the switch chip to determine which output port to route the packet to based on the route nibble. Multicast packets give the switch the ability to replicate and distribute packets to predefined groups of nodes.

7.2.2 SP Switch (F/C 4011)

The SP Switch provides low latency, high-bandwidth communication between nodes supplying a minimum of four paths between any pair of nodes. The SP Switch can be used in conjunction with the SP Switch Router to dramatically speed up TCP/IP, file transfers, remote procedure calls, and relational database functions. The SP Switch offers the following improvements over the High Performance series of switches:

- Higher availability
- Fault isolation
- Concurrent maintenance for nodes
- Improved switch chip bandwidth

The required SP Switch adapter (F/C 4020), SP Switch MX2 adapter (F/C 4023), or the withdrawn SP Switch MX adapter (F/C 4022) connects each SP node to the SP Switch subsystem. One adapter of the required type must be ordered for each node in a switch configured SP system. If you are using a switch expansion frame, the SP Switch subsystem will allow you to scale your SP system up to 128 nodes.

When you order F/C 4011, you will receive one 16-port SP Switch and all of the internal cables needed to connect up to sixteen nodes to the switch. Internal cables can only be used in the frame that the switch is mounted in. If you are connecting nodes mounted in a non-switched expansion frame to a switch, you must use separately ordered external cables. All frame-to-frame cables needed to make switch connections between frames must also be ordered separately.

7.2.3 SP Switch-8 (F/C 4008)

Eight port switches are a low cost alternative to the full size 16 port switches. The 8-port SP Switch-8 (SPS-8) (F/C 4008) provides switch functions for up to eight processor nodes in Model 500 and Model 550 systems. The SP Switch-8 is compatible with high nodes.

When you order F/C 4008, you will receive one 8-port SP Switch and all of the internal cables needed to connect up to eight nodes to the switch. Internal cables can only be used in the frame that the switch is mounted in. If you are connecting nodes mounted in a non-switched expansion frame to a switch, you must use separately ordered external cables.

An SP Switch-8 can be configured in one of two ways:

1. In a Model 500 (1.25 m) frame with up to four F/C 1500 (1.25 m) non-switched expansion frames attached
2. In a Model 550 (1.93 m) frame with F/C 1550 non-switched expansion frames supporting up to eight nodes

The SP Switch-8 has two active switch chip entry points. Therefore, your ability to create system partitions is restricted with this switch. With the maximum eight nodes attached to the switch, you have two possible system configurations:

1. A single partition eight node system.
2. Two system partitions with four nodes each.

For upgrades to greater than eight node support, F/C 4008 is replaced by the SP Switch F/C 4011. The SP Switch-8 uses a similar network topology, proprietary protocol, and communication physical layer as F/C 4011.

The SP-attached server cannot be attached to the SP Switch-8.

7.2.4 SP Switch2 adapter (F/C 4025)

The SP Switch2 adapter encompasses many new hardware changes. They include:

- Up to two adapters, each adapter contains two switch ports, can be placed in a single POWER3 SMP High node
- Faster interface, data paths, and microprocessor
- New data memory component

Using multiple ports and adapters in a single node, along with the new SP Switch2 adapter, will increase the overall bandwidth of the switch adapter and decrease the latency of switch communications on the POWER3 SMP High node.

Hardware changes

The design of some chips are changed and a new chip is also added. This section describes the new SP Switch2 chip design.

Figure 26 shows the SP Switch MX Adapter Node.

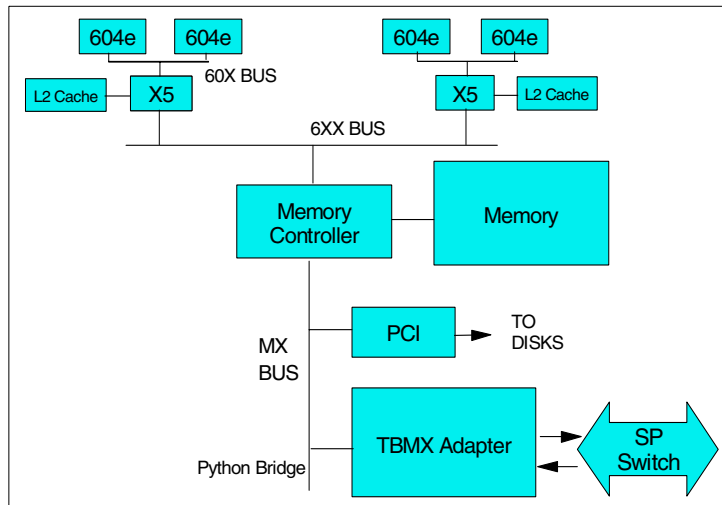


Figure 26. SP Switch MX Adapter with 332 MHz SMP Node

Figure 27 shows SP Switch2 Adapter Hardware Structure.

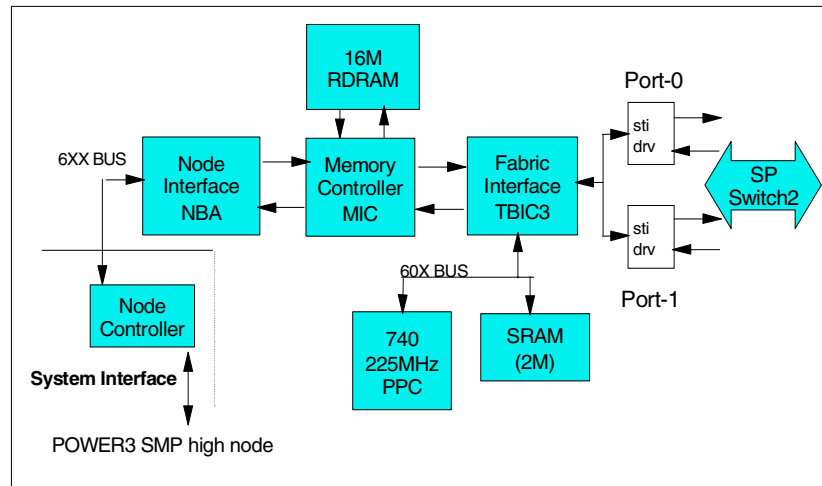


Figure 27. SP Switch2 Adapter Hardware Structure

The SP Switch2 adapter connects the system interface directly. This design solves the performance bottleneck caused by the bus contention, and enables switch traffic to proceed at higher bandwidths and lower latencies.

The TBIC3 chip moves data from the switch chip onto the adapter. Its function is very similar to those provided by TBIC on the TB3 adapter or the TBIC2 on the TB3MX adapter. The primary enhancements are its faster data transfer path (500 MB/s) and the addition of packet reassembly hardware in the chip.

The NBA chip moves data between the adapter and POWER3 SMP node's 6xx bus. The NBA chip's function is similar to that provided by the MBA, which was supplied on the TB3MX. The primary change is that it gives the adapter a memory type bus interface rather than the I/O style used in previous adapters.

The MIC chip is responsible for either passing information between the TBIC3 and NBA, or for giving access to the adapter's data memory component. This new function not available on previous adapters.

The 740 microprocessor is responsible for passing control information between the adapter and the software on the POWER3 SMP node, for encoding or formatting packet headers, for building packet routing information, and for handling error conditions.

7.2.5 SP Switch adapter (F/C 4020, 4022, and 4023)

If you plan to use a switch in your SP system, you will need a switch adapter to connect each RS/6000 SP node to the switch subsystem. SP Switches use either the SP Switch adapter (F/C 4020), the SP Switch MX2 adapter (F/C 4023), or the withdrawn SP Switch MX adapter (F/C 4022). The SP Switch adapter is used in MCA type nodes, while the SP Switch MX or MX2 adapters are used in PCI type nodes. One switch adapter is needed for each node in the system.

Attention

- The SP Switch MX adapter (F/C 4022) has been withdrawn from production. This adapter has been replaced with the SP Switch MX2 adapter (F/C 4023).
- High Performance Switch adapters are not compatible with any of the SP Switch adapters. These adapters cannot coexist in the same system configuration.

Table 15 shows the SP Switch adapter features.

Table 15. SP Switch adapter features

F/C	Description
4020	SP Switch adapter for installation as follows: - Optional - Order one adapter per MCA node
4022	SP Switch MX adapter (SP switch adapter for 332 MHz SMP nodes) for installation as follows: - Optional - Order one adapter per PCI node - Withdrawn 10/98
4023	SP Switch MX2 adapter (SP switch adapter for POWER3 SMP nodes and 332 MHz SMP nodes) for installation as follows: - Optional - Order one adapter per PCI node
4018	HiPS adapter-2(s) for installation as follows: - Withdrawn, only available for existing systems - Order one adapter per node
4017	HiPS adapter-1(s) for installation as follows: - Withdrawn

7.2.6 Special SP Switch adapter (F/C 4021 and F/C 8396)

Some optional SP system equipment requires special adapters in order to connect to the SP Switch network. These devices are:

- SP Switch Router (Machine Type 9077)
- SP-attached server (Machine Type 7017)

7.2.6.1 SP Switch Router (Machine Type 9077)

- Requires one SP Switch Router adapter (F/C 4021) for each SP Switch connection. This adapter is ordered with the SP Switch Router.
- The SP Switch Router adapter is placed in the SP Switch Router. The included cable attaches to the SP Switch and uses up one valid node slot on the switch.

7.2.6.2 SP-attached server (Machine Type 7017)

- Requires one RS/6000 SP System Attachment adapter (RS/6000 F/C 8396) only if the SP-attached server is mounted in a switch-configured system. This adapter is ordered with the Enterprise Server.

- The RS/6000 SP System Attachment adapter is placed in the SP-attached server and requires a cable to connect with the SP Switch. Requires one valid switch port on the SP Switch. The cable is ordered with the SP system.
- SP-attached servers do not require this adapter if used in a non-switched SP system.
- This adapter is not compatible with the SPS-8 switch.
- Only one adapter is allowed to be installed in each SP-attached server.

Table 16 shows the special SP Switch adapter features.

Table 16. Special SP Switch adapter features

F/C	Description
4021	SP Switch Router adapter for installation as follows: - Required for the RS/6000 Switch Router - Multiple adapters allowed in each SP Switch Router
RS/6000 F/C 8396	RS/6000 SP System Attachment adapter (SP Switch adapter used in SP-attached servers) for installation as follows: - Required in servers attached to a switch-configured SP system - Only one adapter is allowed per server

Chapter 8. Control workstations

The RS/6000 SP system requires a customer-supplied RS/6000 workstation with a color monitor. The control workstation serves as a point of control for managing, monitoring, and maintaining the RS/6000 SP frames and individual processor nodes. A system administrator can perform these control tasks by logging into the control workstation from any other workstation on the network.

The control workstation also acts as a boot/install server for other servers in the RS/6000 SP system. In addition, the control workstation can be set up as an authentication server using Kerberos. The control workstation can be the Kerberos primary server, with the master database and administration service, as well as the ticket-granting service. As an alternative, the control workstation can be set up as a Kerberos secondary server, with a backup database, to perform ticket-granting service.

Kerberos is no longer the only security method. The Distributed Computing Environment (DCE) can be used with Kerberos V4, or by itself, or nothing (AIX std security).

8.1 Overview

There are two basic types of control workstations:

- Control workstations using PCI adapters
- Control workstations using MCA adapters

Both types of control workstations must be connected to each frame through an RS-232 cable and the SP Ethernet shown in Figure 28 on page 124. These 15 meter (50 foot) cables are supplied with each frame. However, the control workstation and SP frames should be no more than 12 meters apart. This leaves three meters of cable for use in the vertical runs between the equipment. If you need longer vertical runs, or if there are under floor obstructions, you must place the components closer together.

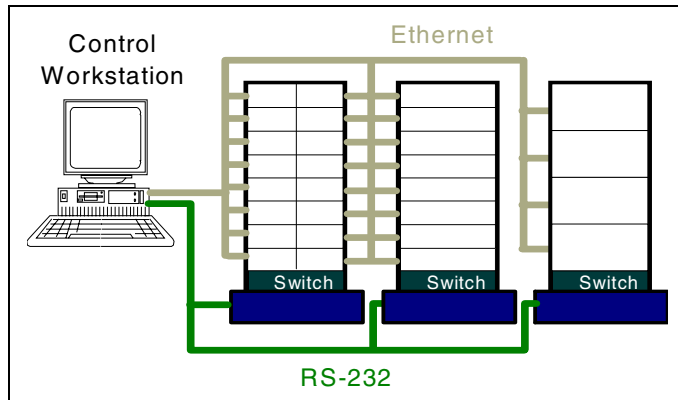


Figure 28. Control workstation interface

All SP-attached servers must also be connected to the control workstation. However, each SP-attached server requires two RS-232 connections as well as the SP Ethernet connection. See Section 4.3.1, “Connecting to the control workstation” on page 86 for details.

Note that most PCI control workstations provide either a 10BaseT or AUI connection for the SP Ethernet LAN. If you are attaching to nodes or SP-attached servers that require a BNC connection, make sure you have ordered the appropriate transceiver.

All SP Switch Routers must also be connected to the control workstation. See Section 6.3.1, “Connecting to the control workstation” on page 105 for details.

8.2 Installation requirements

The control workstation and some of its software are not part of the SP package and must be ordered separately. Make sure you have ordered them in time so they will arrive when the rest of your SP does. To coordinate delivery of the SP and control workstation, your IBM representative should link the SP and control workstation orders with a System Order Number. These requirements are in the following categories:

- Supported RS/6000 workstations
- System requirements
- Interface adapter requirements
- Software requirements

8.2.1 Supported RS/6000 workstations

Supported IBM RS/6000 workstations are listed for two types separately:

- PCI RS/6000 workstations
- MCA RS/6000 workstations

8.2.1.1 PCI RS/6000 workstations

Table 17 shows supported RS/6000 workstations as PCI control workstations:

Table 17. Supported RS/6000 PCI control workstations

Machine Type	Model
RS/6000 7024	E20 and E30 ¹
RS/6000 7025	F30 ^{1,2}
RS/6000 7025	F40 and F50 ^{3,4}
RS/6000 7026	H10 and H50 ^{3,4}
RS/6000 7043	140 and 240 ^{3,5}

¹Supported by PSSP 2.2 and later

²On systems introduced since PSSP 2.4, either the 8-port (F/C 2493) or 128-port (F/C 2944) PCI bus asynchronous adapter should be used for frame controller connections. IBM strongly suggests you use the support processor option (F/C 1001). If you use this option, the frames must be connected to a serial port on an asynchronous adapter and not to the serial port on the control workstation planar board.

³The native RS-232 ports on the system planar can not be used as tty ports for the hardware controller interface. The 8-port asynchronous adapter EIA-232/ RS-422, PCI bus (F/C 2943), or the 128-port Asynchronous Controller (F/C 2944) are the only RS232 adapters that are supported. These adapters require AIX 4.2.1 or AIX 4.3 on the control workstation.

⁴IBM strongly suggests you use the support processor option (#1001).

⁵The 7043 can only be used on SP systems with up to four frames. This limitation applies to the number of frames and not the number of nodes. This number includes expansion frames.

8.2.1.2 MCA RS/6000 workstations

Table 18 shows supported RS/6000 workstations as MCA control workstations.

Table 18. Supported RS/6000 MCA control workstations

Machine Type	Model
RS/6000 7012	37T, 370, 375, 380, 39H, 390, 397, G30, and G40
RS/6000 7013	570, 58H, 580, 59H, 590, 591, 595, J30, J40, and J50 ¹
RS/6000 7015	97B, 970, 98B, 980, 990, R30, R40, and R50 ^{1,2}
RS/6000 7030	3AT, 3BT, 3CT
¹ Requires a 7010 Model 150 X-Station and display. Other models and manufacturers that meet or exceed this model can be used. An ASCII terminal is required as the console.	
² Installed in either the 7015-99X or 7015-R00 rack.	

8.2.2 System requirements

The minimum requirements for the control workstation are:

- At least 128 MB of main memory. An extra 64 MB of memory should be added for each additional system partition. For SP systems with more than 80 nodes, 256 MB is required, and 512 MB of memory is recommended.
- 4 GB of disk storage. If the SP is going to use an HACWS configuration, you can configure 2 GB of disk storage in the rootvg volume group and 2 GB in an external volume group.

Because the control workstation is used as a Network Installation Manager (NIM) server, the number of unique file sets required for all the nodes in the SP system might be larger than a normal single system. You should plan to reserve 2 GB of disk storage for the file sets and 2 GB for the operating system. This will allow adequate space for future maintenance, system mksysb images, and LPP growth. Keep in mind that if you have nodes at different levels of PSSP or AIX, each node requires its own LPP source, which will take up extra space.

A good rule of thumb to use for disk planning for a production system is 4 GB for the rootvg to accommodate additional logging and /tmp space plus 1 GB for each AIX release and modification level for lppsource files. Additional disk space should be added for mksysb images for the nodes.

If you plan on using rootvg mirroring, then double the number of physical disks you estimated so far.

- Physically installed with the RS-232 cable to within 12 meters of each SP frame.
- Physically installed with two RS-232 cables to within 12 meters of each SP-attached server, such as an RS/6000 Enterprise Server Model S70, S70 Advanced or S80.
- Equipped with the following I/O devices and adapters:
 - A 3.5 inch diskette drive.
 - Four or eight millimeter (or equivalent) tape drive.
 - A SCSI CD-ROM device.
 - One RS-232 port for each SP frame.
 - Keyboard and mouse.
 - Color graphics adapter and color monitor. An X-station model 150 and display are required if an RS/6000 that does not support a color graphics adapter is used.
 - An appropriate network adapter for your external communication network. The adapter does not have to be on the control workstation. If it is not on the control workstation, the SP Ethernet must extend to another host that is not part of the SP system. A backup control workstation does not satisfy this requirement. This additional connection is used to access the control workstation from the network when the SP nodes are down.
 - SP Ethernet adapters for connection to the SP Ethernet

The number of Ethernet adapters required depends completely on the Ethernet topology you use on your SP system. The following types of Ethernet adapters can be used:

- Ethernet adapters with thin BNC.

Each Ethernet adapter of this type can have only 30 network stations on a given Ethernet cable. The control workstation and any routers are included in the 30 stations.
- Ethernet adapters with twisted pair (RJ45/AUI). A network hub or switch is required.
- 10/100 Mbps Ethernet adapters. A network hub or switch is required.

8.2.3 Interface adapter requirements

Several different control workstations are available. Each model has different communications adapters offered as standard equipment. Depending on the model you choose, serial and Ethernet adapters may have to be added to the workstation to satisfy the needs of your SP system.

8.2.3.1 Serial port adapters

Supported serial adapters are listed for two types separately:

- PCI control workstations
- MCA control workstations

PCI control workstations

All new PCI control workstations require a minimum of one additional asynchronous adapter. For additional PCI serial ports, select the equipment you need from the feature codes shown in Table 19.

Table 19. Serial port adapters for PCI control workstations

F/C	Description
8-port	
2931	8-port asynchronous adapter ISA BUS EIA-232 (withdrawn 12/97)
2932	8-port asynchronous adapter ISA BUS EIA-232/422A (withdrawn 12/97)
2934	8-port asynchronous adapter PCI BUS EIA-232/RS-422
128-port	
2933	128-port asynchronous controller ISA BUS (withdrawn 12/97)
2944	128-port asynchronous controller PCI BUS
8130	1.2 MBps Remote Asynchronous Node (RAN) 16-port EIA-232 (US)
8131	128-port asynchronous controller cable, 4.5 m (1.2 MBps transfers)
8132	128-port asynchronous controller cable, 23 cm (1.2 MBps transfers)
8133	RJ-45 to DB-25 converter cable
8134	World trade version of F/C 8130
8136	1.2 MBps Rack mountable Remote Asynchronous Node (RAN) 16-port EIA-232
8137	2.4 MBps Enhanced Remote Asynchronous Node (RAN) 16-port EIA-232
8138	2.4 MBps Enhanced Remote Asynchronous Node (RAN) 16-port RS-422

F/C	Description
2934	Asynchronous Terminal/Printer Cable, EIA-232 (2.4 MBps transfers)
3124	Serial port to serial port cable for drawer-to-drawer connections (2.4 MBps transfers)
3125	Serial port to serial port cable for rack-to-rack connections (2.4 MBps transfers)

In addition to the listed PCI bus adapters, the 7024-EXX and 7025-F30 control workstations will also support the listed ISA bus adapters. All other PCI control workstations will only support PCI bus adapters. PCI adapters offer performance advantages in all PCI control workstations and should be used whenever possible.

MCA control workstations

For additional MCA serial ports, select the equipment you need from the feature codes shown in Table 20.

Table 20. Serial port adapters for MCA control workstations

F/C	Description
8-port	
2930	8-port asynchronous adapter
2995	Multiport interface cable
16-port	
2955	16-port asynchronous adapter (F/C 2955 is not compatible with the SP-attached server.)
2996	Multiport interface cable
128-port	
8128	128-port asynchronous controller
8130	Remote asynchronous node 16-port EIA-232
8134	World trade version of F/C 8130

8.2.3.2 Ethernet adapters

Supported Ethernet adapters are listed for two types separately:

- PCI control workstations
- MCA control workstations

PCI control workstations

For additional PCI Ethernet ports, select the equipment you need from the feature codes shown in Table 21.

Table 21. Ethernet adapters for PCI control workstations

F/C	Description
2968	IBM 10/100 Mbps Ethernet PCI adapter
2985	PCI Ethernet BNC/RJ-45 adapter
2987	PCI Ethernet AUI/RJ-45 adapter
4224	Ethernet 10Base2 transceiver

MCA control workstations

For additional MCA Ethernet adapters, select the equipment you need from the feature codes shown in Table 22.

Table 22. Ethernet adapters for MCA control workstations

F/C	Description
2980	Ethernet high performance LAN adapter
2922	Ethernet Twisted Pair (TP) adapter
2993	Ethernet BNC/AUI adapter
4224	Ethernet 10Base2 transceiver

8.2.4 Software requirements

The control workstation requires the following software:

- AIX 4.3.3 (or later) server (5765-C34)
- PSSP 3.2 (5765-D51)
- C for AIX 3.6.6 (or later) or VisualAge C++ Professional 4.0 for AIX (or later)

At least one concurrent use license is required for the SP system. Concurrent licensing is recommended so the one license can float across the SP nodes and the control workstation. It is needed for crash to work effectively and to obtain IBM software support for the SP system. You can order the license as part of the SP system. It is not specifically required on the control workstation if a license server for AIX for C and C++ exists some place in the network and the SP is included in the license server's cell.

8.3 High Availability Control Workstation

The High Availability Control Workstation (HACWS) is a major component of the effort to reduce the possibility of single point of failure opportunities in the SP. There are already redundant power supplies and replaceable nodes. However, there are also many elements of hardware and software that could fail on a control workstation. With a HACWS, your SP system will have the added security of a backup control workstation. Also, HACWS allows your control workstation to be powered down for maintenance or updating without affecting the entire SP system.

The design of the HACWS is modeled on the High Availability Cluster Multi-Processing for the AIX (HACMP) licensed program product. HACWS utilizes HACMP running on two RS/6000 control workstations in a two-node rotating configuration. HACWS utilizes an external disk that is accessed non-concurrently between the two control workstations for storage of SP related data. There is also a dual RS-232 frame supervisor card with a connection from each control workstation to each SP frame in your configuration. This HACWS configuration provides automated detection, notification, and recovery of control workstation failures.

SP-attached servers can be used in your SP system with HACWS, but there is no dual RS232 cabling support for them. See Section 8.3.1.1, "Limits and restrictions" on page 133.

8.3.1 Overview

The SP system looks similar except that there are two control workstations connected to the SP Ethernet and TTY network. The frame supervisor TTY network is modified to add a standby link. The second control workstation is the backup.

Figure 29 shows a logical view of a HACWS. The figure shows disk mirroring, which is an important part of high availability planning.

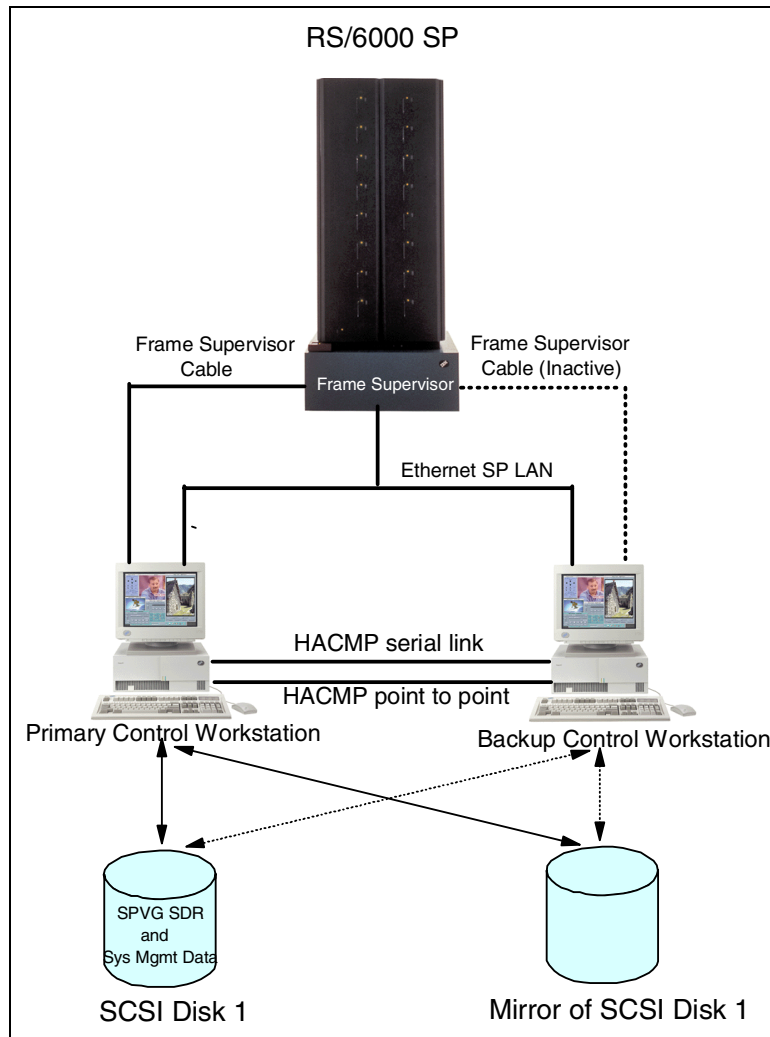


Figure 29. High Availability Control Workstation with disk mirroring

If the primary control workstation fails, there is a disruptive failover that switches the external disk storage, performs IP and hardware address takeover, restarts the control workstation applications, remounts file systems, resumes hardware monitoring, and lets clients reconnect to obtain services or to update control workstation data. This means that there is only one active control workstation at any time.

The primary and backup control workstations are also connected on a private point-to-point network and a serial TTY link or target mode SCSI. The backup control workstation assumes the IP address, IP aliases, and hardware address of the primary control workstation. This lets client applications run without changes. The client application, however, must initiate reconnects when a network connection fails.

The SP data is stored in a separate volume group on the external disk storage.

The backup control workstation can run other unrelated applications if desired. However, if the application on the backup control workstation takes significant resources, that application may have to be stopped during failover and reintegration periods.

8.3.1.1 Limits and restrictions

The HACWS support has the following limitations and restrictions:

- You cannot split the load across a primary and backup control workstation. Either the primary or the backup provides all the functions at one time.
- The primary and backup control workstations must each be an RS/6000. You cannot use a node at your SP as a backup control workstation.
- The backup control workstation cannot be used as the control workstation for another SP system.
- The backup control workstation cannot be a shared backup of two primary control workstations.

There is a one-to-one relationship of primary to backup control workstations; a single primary and backup control workstation combination can be used to control only one SP system.

- If your primary control workstation is a PSSP Kerberos V4 authentication server, the backup control workstation must be a secondary authentication server.
- If you plan to have DCE authentication enabled, you cannot use HACWS. If you already use HACWS, do not enable DCE authentication.
- HACWS does not tolerate IPV6 aliases for IPV4 addresses.

The following apply if you use SP-attached servers or Clustered Enterprise Servers and HACWS support:

- The S70, S70 Advanced, and S80 SP-attached servers are directly attached to the control workstation through two RS-232 serial connections. There is no dual RS-232 hardware support for these

connections like there is for SP frames. These servers can only be attached to one control workstation at a time. Therefore, when a control workstation fails or scheduled downtime occurs, and the backup control workstation becomes active, you will lose hardware monitoring and control and serial terminal support for your SP-attached servers. The specific functions that are lost include:

- Power on and off control.
- Reboot control.
- Serial port communications for s1term.
- Nodecond support to obtain the hardware Ethernet address and to network boot the node.
- Monitoring of the following Hardmon variables and state data, whether using the SP Perspectives graphical user interface, commands (such as `hmmon`, `spmon`, `sphardware`), or RSCT resource variables:
 - `diagByte`
 - `hardwareStatus`
 - `lcd1`
 - `lcd2`
 - `LCDhasMessage`
 - `nodefail1`
 - `nodeLinkOpen1`
 - `nodepower`
 - `serialLinkOpen`
 - `spcn`
 - `SPCNhasMessage`
 - `src`
 - `SRChasMessage`
 - `timeTicks`
- The ability to make configuration changes related to the SP-attached servers. For example, you cannot add new SP-attached servers when the backup is the primary control workstation.
- You can use PSSP to shut down or restart the SP system, but it will not affect SP-attached servers.
- You cannot use PSSP to shutdown or restart a system of clustered enterprise servers.
- The SP-attached servers will have the SP Ethernet connection from the backup control workstation; so, PSSP components requiring this connection will still work correctly. This includes components, such as the

availability subsystems, user management, logging, authentication, the SDR, file collections, accounting, and others.

For a list of the monitoring and control functions that are lost, see the discussion of SP-attached and clustered enterprise servers in the HACWS chapter of the book *PSSP: Administration Guide*, SA22-7348.

8.3.2 Installation requirements

There are a couple of requirements for hardware and software that must be met before you install the HACWS. These requirements are in the following categories:

- System requirements
- Software requirements

8.3.2.1 System requirements

In addition to the system requirements described in Section 8.2.2, “System requirements” on page 126, the following are required:

- Two supported RS/6000 workstations.

Each of these RS/6000s must have the same set of I/O required for control workstations as described in Section 8.2.1, “Supported RS/6000 workstations” on page 125. They can be different models, but the tty configuration must be exactly the same on each control workstation. The disks should be of the same type and configured the same way on both control workstations to allow the hdiskx numbers to be consistent between the two control workstations.

- External disk storage that is supported by HACMP and the control workstation being used.
 - Two external disk controllers and mirrored disks are strongly recommended but not required. If a single external disk controller is used, the control workstation single point of failure has not been eliminated but moved to the disk subsystem.
- The HACWS connectivity feature (F/C 1245) on each SP frame.
- An additional RS232 connection for HACMP communication is needed if target mode SCSI is not being used for the HACMP communication.

8.3.2.2 Software requirements

The software requirements for the control workstation include:

- Two AIX server licenses

- Two licenses for IBM C for AIX 3.6.6 or later or the batch C and C++ 3.6.6 or later compilers and runtime libraries of VisualAge C++ Professional 4.0 for AIX or later

If the compiler's license server is on the control workstation, the backup control workstation should also have a license server with at least one license. If there is no license server on the backup control workstation, an outage on the primary control workstation will not allow the SP system access to a compiler license.

- Two licenses and software sets for High Availability Cluster Multi-Processing for AIX (HACMP)

This is the high availability feature of HACMP. Both the client and server options must be installed on both control workstations. You must purchase two licenses. Do not use the Enhanced Scalability feature.

- PSSP 3.2 optional component HACWS

This is the customization software that is required for HACMP support of the control workstation. It comes with your order of PSSP 3.2 as an optionally installable component. Install a copy on both control workstations.

Chapter 9. Communication adapters

This chapter provides information for available communication adapters for the following processor nodes and dependent node:

- PCI nodes (POWER3 SMP nodes and 332 MHz SMP nodes)
- SP-attached servers
- SP Switch Routers

9.1 PCI nodes communication adapters

This section contains information about supported communication adapters for PCI nodes.

Table 23 shows a list of supported communication adapters. The following sections describe each adapters in detail.

Table 23. Supported communication adapters for PCI nodes

F/C	PCI adapter name
2741	FDDI SK-NET LP SAS
2742	FDDI SK-NET LP DAS
2743	FDDI SK-NET UP DAS
2751	S/390 ESCON Channel Adapter
2920	Token Ring Auto Lanstream
2943	EIA 232/RS-422 8-port Asynchronous Adapter
2944	WAN RS232 128-port
2947	IBM ARTIC960Hx 4-Port Selectable Adapter
2962	2-port Multiprotocol X.25 Adapter
2963	ATM 155 TURBOWAYS UTP Adapter
2968	Ethernet 10/100 MB
2969	Gigabit Ethernet - SX
2985	Ethernet 10 MB BNC
2987	Ethernet 10 MB AUI
2988	ATM 155 MMF

F/C	PCI adapter name
6206	Ultra SCSI Single Ended
6207	Ultra SCSI Differential
6208	SCSI-2 F/W Single-Ended
6209	SCSI-2 F/W Differential
6215	SSA RAID 5
6222	SSA Fast-Write Cache module
6310	IBM ARTIC960RxD Quad Digital Trunk Adapter

9.1.1 FDDI SK-NET LP SAS (F/C 2741)

The SYSKONNECT SK-NET FDDI-LP SAS PCI adapter (F/C 2741) is a fiber optical FDDI Single Attach Station (SAS) that is compatible with the FDDI-ANSI X3T12 specifications and FDDI Standard Series. The adapter provides single attachment to an FDDI concentrator (or point-to-point) using fiber optic cabling (not supplied with the adapter).

Feature characteristics

This feature has the following characteristics:

- Supports single-ring FDDI attachment at 100 Mbps via a customer-supplied FDDI concentrator
- Supports all TCP/IP protocols and ANSI Station Management (SMT) 7.3

Feature components

This feature order provides the following:

- Adapter card
- Diagnostic wrap plug
- Diskette with adapter device driver
- Installation instructions

Customer components

You must supply the following components for this feature:

- An FDDI concentrator, such as the IBM 8240 (or equivalent) concentrator, to connect to your FDDI local area network
- One 62.5/125 micron multimode fiber duplex cable with SC connectors

Plugging rules

This feature has the following plugging rules:

- A maximum of two adapters per POWER3 SMP thin node.
- A maximum of four adapters per POWER3 SMP wide node.
- A maximum of two adapters per 332 MHz SMP thin node.
- A maximum of four adapters per 332 MHz SMP wide node.
- The sum of the following adapters may be used in any combination but must not exceed four total per SMP wide node: F/C 2741, 2742, 2743, 2963, 2968, and 2988.
- This adapter should not be used in slots I5, I6, I7, or I8 in 332 MHz SMP wide nodes. Performance of this adapter is reduced in those slots.
- A configuration using more than one SysKconnect FDDI adapter (F/C 2741, 2742, or 2743) where any one of them is F/C 2743 (SysKconnect SK-NET FDDI-UP SAS PCI) constitutes a Class A system.

Hardware requirement

This feature has the following hardware requirement:

- One PCI adapter slot

Software requirements

This feature has the following software requirements:

- AIX 4.2.1, or later, installed on the node
- PSSP 2.4, or later, installed on the node
- Adapter device driver and FDDI common code (provided with adapter)

9.1.2 FDDI SK-NET LP DAS (F/C 2742)

The SYSKCONNECT SK-NET FDDI-LP DAS PCI adapter (F/C 2742) is a fiber optical FDDI Dual Attach Station (DAS) that is compatible with the FDDI-ANSI X3T12 specifications and FDDI Standard Series. The adapter provides either dual attachment to the main ring path or dual homing to one or two FDDI concentrators using fiber optic cabling (not supplied with the adapter).

Feature characteristics

This feature has the following characteristics:

- Supports dual ring FDDI attachment at 100 Mbps
- Supports all TCP/IP protocols and ANSI Station Management (SMT) 7.3

Feature components

This feature order provides the following:

- Adapter card
- Diagnostic wrap plug
- Diskette with adapter device driver
- Installation instructions

Customer components

You must supply the following components for this feature:

- A FDDI concentrator, such as the IBM 8240 (or equivalent) concentrator, to connect to the FDDI network for dual homing configurations
- Two 62.5/125 micron multimode fiber duplex cables with SC connectors

Plugging rules

This feature has the following plugging rules:

- A maximum of two adapters per POWER3 SMP thin node.
- A maximum of four adapters per POWER3 SMP wide node.
- A maximum of two adapters per 332 MHz SMP thin node.
- A maximum of four adapters per 332 MHz SMP wide node.
- The sum of the following adapters may be used in any combination but must not exceed four total per SMP wide node: F/C 2741, 2742, 2743, 2963, 2968, and 2988.
- Should not be used in slots I5, I6, I7, or I8 in 332 MHz SMP wide nodes. Performance of this adapter is reduced in those slots.
- A configuration using more than one SysKonnnect FDDI adapter (F/C 2741, 2742, or 2743) where any one of them is F/C 2743 (SysKonnnect SK-NET FDDI-UP SAS PCI) constitutes a Class A system.

Hardware requirement

This feature has the following hardware requirement:

- One PCI adapter slot

Software requirements

This feature has the following software requirements:

- AIX 4.2.1, or later, installed on the node
- PSSP 2.4, or later, installed on the node
- Adapter device driver and FDDI common code (provided with adapter)

9.1.3 FDDI SK-NET UP DAS (F/C 2743)

The SYSKONNECT SK-NET FDDI-UP DAS PCI adapter (F/C 2743) is a fiber optical FDDI Dual Attach Station (DAS) that is compatible with the FDDI-ANSI X3T12 specifications and FDDI Standard Series. The adapter provides single attachment to a FDDI concentrator (or point to point) using Category 5 Unshielded Twisted Pair (UTP) cabling (not supplied with the adapter).

Feature characteristics

This feature has the following characteristics:

- Supports single ring FDDI attachment at 100 Mbps
- Supports all TCP/IP protocols and ANSI Station Management (SMT) 7.3

Feature components

This feature order provides the following:

- Adapter card
- Diagnostic wrap plug
- Diskette with adapter device driver
- Installation instructions

Customer components

You must supply the following components for this feature:

- An FDDI concentrator, such as the IBM 8240 (or equivalent) concentrator, to connect to the FDDI network for dual homing configurations
- One UTP Category 5 cable

Plugging rules

This feature has the following plugging rules:

- A maximum of two adapters per SMP thin node.
- A maximum of four adapters per SMP wide node.
- The sum of the following adapters may be used in any combination, but must not exceed six total per POWER3 SMP wide node: F/C 2741, 2742, 2743, 2963, 2968, and 2988.
- The sum of the following adapters may be used in any combination, but must not exceed four total per 332 MHz SMP wide node: F/C 2741, 2742, 2743, 2963, 2968, and 2988.
- The adapter should not be used in slots I5, I6, I7, or I8 in 332 MHz SMP wide nodes. Performance of this adapter is reduced in those slots.

- A configuration using more than one SysKonnnect FDDI adapter (F/C 2741, 2742, or 2743) where any one of them is F/C 2743 (SysKonnnect SK-NET FDDI-UP SAS PCI), which constitutes a Class A system.

Hardware requirement

This feature has the following hardware requirement:

- One PCI adapter slot

Software requirements

This feature has the following software requirements:

- AIX 4.2.1, or later, installed on the node
- PSSP 2.4, or later, installed on the node
- Adapter device driver and FDDI common code (provided with adapter)

9.1.4 S/390 ESCON Channel adapter (F/C 2751)

The PCI S/390 ESCON Channel adapter (F/C 2751) provides the SP system with an attachment to IBM Enterprise Systems Connection (ESCON) channels on System/390 mainframes. This direct ESCON channel connection provides a fiber optic link that can take advantage of ESCON Directors (fiber optic switches) permitting multiple channel connections. The adapter supports: VM/ESA, MVS/ESA, and OS/390.

Feature characteristics

This feature has the following characteristics:

- Full length PCI adapter
- Supports attachment to either 10 MB or 17 MB ESCON channels
- Supports VM, MVS, and OS/390
- Supports CLIO/S
- Supports ESCON multiple Image Facility (EMIF)
- Maximum distance supported, 43 Km using LED and XDF ESCON links
- S/390 TCP/IP for VM and MVS
- PCI 32-bit Bus Master adapter

Feature components

This feature order provides the following:

- One full length PCI adapter
- CD-ROM with device drivers

- Instruction manual
- Diagnostic wrap plug

Customer components

The customer must supply the following components for this feature:

- ESCON cabling, requires 62.5/125 multimode fiber cable with ESCON duplex connectors on both ends
- AIX program feature, ESCON Control Unit LPP (5765-D49)

Plugging rules

This feature has the following plugging rules:

- A maximum of one adapter per SMP thin node.
- A maximum of two adapters per SMP wide node.
- If two ESCON adapters are used in an SMP wide node, one adapter must be placed in slot I1 or I2 on the CPU side. The other adapter must be placed in either slot I1, I2, I3, or I4 of the first I/O bus.

The S/390 ESCON adapter cannot be plugged into slot I5, I6, I7, or I8 in SMP wide nodes.

Hardware requirement

This feature has the following hardware requirement:

- One PCI adapter slot

Software requirements

This feature has the following software requirements:

- AIX 4.3.2 or later
- PSSP 3.1 or later
- Device drivers (included with adapter)
- ESCON Control Unit LPP (separately ordered as LPP 5765-D49)

9.1.5 Token Ring Auto LANstreamer (F/C 2920)

The PCI Auto LANstreamer Token Ring adapter (F/C 2920) is a PCI 16/4 Token Ring adapter that is compatible with IEEE 802.5 specifications. The adapter has two external connections: RJ-45 to attach to UTP cabling and a 9-pin D-Shell to attach to STP cabling.

Feature characteristics

This feature has the following characteristics:

- Complies with IEEE 802.5 specifications
- Attaches to 4 MBps or 16 MBps token-ring area networks
- Supports both full and half duplex operations
- Has a PCI 32-bit Bus Master adapter

Feature components

This feature order provides the following:

- Adapter card
- Diskette with adapter device driver
- Installation instructions

Customer components

The customer must supply the following components for this feature:

- Network equipment, such as a MAU and/or switching hub, to connect the Token Ring network
- UTP or STP cable to attach the adapter to the Token Ring network

Plugging rules

This feature has the following plugging rules:

- A maximum of two adapters per POWER3 SMP thin node
- A maximum of eight adapters per POWER3 SMP wide node
- A maximum of two adapters per 332 MHz SMP thin node
- A maximum of nine adapters per 332 MHz SMP wide node

Hardware requirement

This feature has the following hardware requirement:

- One PCI adapter slot

Software requirements

This feature has the following software requirements:

- AIX 4.2.1, or later, installed on the node
- PSSP 2.4, or later, installed on the node
- Adapter device driver

9.1.6 EIA 232/RS-422 8-Port Asynchronous adapter (F/C 2943)

The 8-port Async feature (F/C 2943) provides the RS/6000 SP with up to eight EIA 232 or RS-422 asynchronous serial lines from a single PCI bus slot.

This adapter adheres to the PCI Revision 2.1 standards for EIA 232 and RS-422. It features a low cost, high performance 32-bit card, 33 MHz bus speed, and a PCI bus transfer rate of 132 MBps.

This adapter provides a single DB-78 output that connects directly to the 8-port DB-25 connector box. All eight ports are software programmable to support either protocol at baud rates up to 230 K. The full set of modem control lines for asynchronous communication are provided for each port. Devices such as terminals, modems, processors, printers, and controllers may be attached.

Feature characteristics

This feature has the following characteristics:

- 8-port asynchronous device connections
- 32-bit Bus Master PCI bus (132 MB per second)
- Short-form factor PCI adapter
- EIA-232 maximum distance 31 m and 62 m dependent on baud rate and RAN
- RS-422 maximum distance 1200 m dependent on baud rate
- 230 K maximum baud rate
- Supports TxD, RxD, RTS, CTS, DSR, DCD, DTR, and RI on EIA 232
- Supports +TxD, -TxD, +RxD, and -RxD on RS-422

Feature components

This feature order provides the following:

- Adapter card
- 25-pin diagnostic wrap plu
- Diskette with adapter device driver
- Installation instructions
- External 3 m DB78 cable to 8-port DB25 breakout box

Customer components

A 3 m cable with attached breakout box is supplied with each adapter. You must supply all cables needed to connect peripheral equipment to this adapter.

Plugging rules

This feature has the following plugging rules:

- A maximum of two adapters per POWER3 SMP thin node.
- A maximum of six adapters per POWER3 SMP wide node.
- A maximum of two adapters per 332 MHz SMP thin node.
- A maximum of six adapters per 332 MHz SMP wide node.
- This adapter may be placed into any of the PCI slots available in either the 332 MHz SMP thin node or 332 MHz SMP wide node.

Hardware requirement

This feature has the following hardware requirement:

- One PCI adapter slot

Software requirements

This feature has the following software requirements:

- AIX 4.3.2, or later, installed on the node
- PSSP 3.1, or later, installed on the node
- Adapter device driver LPP Image (provided with adapter)

9.1.7 WAN RS232 128-port (F/C 2944)

The 128-port Async feature (F/C 2944) provides the RS/6000 SP with up to 128 EIA-232 asynchronous serial lines from a single PCI bus slot. This adapter adheres to the PCI standard. It features a low cost, high performance 32-bit card, 33 MHz bus speed, and a PCI bus transfer rate of 132 MBps.

Two 2.4 MBps synchronous channels link the adapter to a maximum of eight 16-port Remote Async Nodes (RANs). Each synchronous channel uses an HD-15 female connector to link up to four RANs. Each RAN supports either EIA-232 or RS-422 connections (16 per RAN), and up to eight RANs may be connected together yielding a total of 128 ports. The RAN utilizes an RJ-45 connector to provide interface signals at speeds up to 230 K baud at a limited number of ports.

Feature characteristics

This feature has the following characteristics:

- 32-bit Bus Master PCI bus
- Two synchronous channels to RAN
- EIA-232 maximum distance of 31 m and 62 m depending on baud rate and RAN
- RS-422 maximum distance 1200 m dependent on baud rate

Plugging rules

This feature has the following plugging rules:

- A maximum of two adapters per SMP thin node.
- A maximum of seven adapters per SMP wide node.
- This adapter may be placed into any of the PCI slots available in the SMP thin or wide node.

Customer components

F/C 2944 utilizes the following optional RANs and device cables, which are available from IBM:

- 1.2 MBps RANs and cables:
 - F/C 8130** 1.2 MBps Remote Asynchronous Node, 16-port, EIA-232 (US)
 - F/C 8131** 128-port Asynchronous Controller Node Cable, 4.5 m
 - F/C 8132** 128-port Asynchronous Controller Cable 23 cm (9 in.)
 - F/C 8133** RJ-45 to DB-25 Converter Cable
 - F/C 8134** 1.2 MBps Remote Asynchronous Node, 16-port, EIA-232 (world trade)
 - F/C 8136** 1.2 MBps Rack Mountable Remote Asynchronous Node, 16-port, EIA-232
- 2.4 MBps RANs and cables:
 - F/C 8137** 2.4 MBps Enhanced Remote Asynchronous Node, 16-port, EIA-232
 - F/C 8138** 2.4 MBps Enhanced Remote Asynchronous Node, 16-port, RS-422
 - F/C 2934** Asynchronous Terminal/Printer Cable, EIA-232
 - F/C 3124** Serial port to serial port cable for drawer-to-drawer connections
 - F/C 3125** Serial port to serial port cable for rack-to-rack connections

Hardware requirement

This feature has the following hardware requirement:

- One PCI adapter slot

Software requirements

This feature has the following software requirements:

- AIX 4.2.1, or later, installed on the node

- PSSP 2.4, or later, installed on the node
- Adapter device driver LPP Image (provided with adapter)

9.1.8 ARTIC960Hx 4-port Selectable adapter (F/C 2947)

The ARTIC960Hx 4-Port Selectable PCI adapter is a one-slot, standard-length, 32-bit PCI card. It provides 4-Ports of either EIA-232, EIA530, RS-449, X.21, or V.35. Only one standard can be used at a time. Each port supports speeds up to 2.0 Mbps. Software support is provided by ARTIC960 Support for AIX, Developer's Kit, AIX versions 4.2.1 or 4.3.2 or later, that provide SDLC and Bisync support. The adapter can also be used for real-time device control, telephony signaling, and custom serial communication protocols.

This adapter is also equipped with a high-performance, eight-channel DMA controller. This DMA controller supports intelligent DMA operations, such as data buffer chaining and end-of-frame processing, to support high-performance communications protocols and high-throughput applications. The DMA controller is fully programmable for OEM and third-party device drivers.

Feature characteristics

This feature has the following characteristics:

- One 120-pin port
- Supports up to four connections of the same type
- Data transfer rates of up to 2 Mbps
- Supported interfaces are:
 - EIA-232
 - EIA-530
 - RS-449
 - X.21
 - V.35
- Support for SDLC and X.25 full-duplex, synchronous protocols

Feature components

This feature order provides the following:

- One ARTIC960Hx adapter (F/C 2947)
- A connecting cable (required). The following are available from IBM:

- F/C 2861** ARTIC960Hx 4-port EIA-232 cable
- F/C 2862** ARTIC960Hx 4-port RS-449 cable
- F/C 2863** ARTIC960Hx 4-port X.21 cable
- F/C 2864** ARTIC960Hx 4-port V.35 (DTE) cable
- F/C 2865** ARTIC960Hx 4-port EIA-530 cable

Plugging rules

This feature has the following plugging rules:

- POWER3 thin nodes
 - A maximum of two adapters per node.
- POWER3 wide nodes
 - A maximum of six adapters per node.
- 332 MHz thin nodes
 - A maximum of two adapters per node.
- 332 MHz wide nodes
 - A maximum of six adapters per node.
 - This adapter is not supported in PCI bus slots I5, I6, I7, or I8 on the I/O side of the 332 MHz wide node.

Hardware requirement

This feature has the following hardware requirement:

- One 32-bit PCI adapter slot

Software requirements

This feature has the following software requirements:

- AIX 4.2.1 and APAR IX81861, AIX 4.3.2 and APAR IX81860 (for SDLC or Bisync), or later
- Adapter device driver (provided with adapter)

9.1.9 2-port Multiprotocol X.25 adapter (F/C 2962)

The 2-Port Multiprotocol adapter (F/C 2962) provides the RS/6000 SP with high speed connections between stand alone system units on a wide area network (WAN). This adapter adheres to the PCI standard and also supports SDLC and X.25 protocols. The 2-port Multiprotocol adapter connects to WAN lines through externally attached data communication equipment including Channel Service Units (CSU), Data Service Units (DSU), and synchronous modems.

This adapter operates at speeds up to 2.048 Mbps and provides two ports that accommodate four selectable interfaces. These interfaces are:

- EIA 232D/V.24
- V.35
- V.36/EIA 449
- X.21

Interface configuration is selected by the type of cable attached. These cables are ordered separately, and you may configure with the 2-Port Multiprotocol adapter with two different cables.

Feature characteristics

This feature has the following characteristics:

- 32-bit Bus Master PCI 2.1 adapter
- Provides two, 36-pin high density (male) ports
- Provides four interface types, EIA 232D/V.24, V.35, V.36/EIA 449, and X.21
- Simultaneously supports two different interfaces
- Supports SDLC and X.25 full duplex synchronous protocols

Plugging rules

This feature has the following plugging rules:

- A maximum of two adapters per POWER3 SMP thin node.
- A maximum of six adapters per POWER3 SMP wide node.
- A maximum of two adapters per 332 MHz SMP thin node.
- A maximum of six adapters per 332 MHz SMP wide node.

With 332 MHz SMP wide nodes, this adapter cannot be plugged into slot 15, 16, 17, or 18.

Customer components

If you plan to operate this adapter using X.25 protocols, then you must separately order the IBM AIXLINK/X.25 LPP (5696-926). This package provides a V.24, V.35, or X.21 port connection to X.25 packet switched networks.

The system interface is determined by the cable connected to this adapter. See Table 24 for a list of available cables and the interface supported by each cable.

The 2-port Multiprotocol adapter can be configured with different cable types on each port.

Table 24. Cable information for 2-port Multiprotocol adapter

F/C	Interface Configuration	Cable Terminations (Length)
2951	EIA 232D/V.24 cable	36-pin to male DB25 (3 m)
2952	V.35 cable	36-pin to 34-pin male (3 m)
2953	V.36/EIA 449 cable	36-pin to 37-pin male (3 m)
2954	X.21 cable	36-pin to male DB15 (3 m)

Hardware requirement

This feature has the following hardware requirement:

- One PCI adapter slot

Software requirements

This feature has the following software requirements:

- AIX 4.3.2, or later.
- PSSP 3.1, or later.
- SDLC protocol support provided as part of the AIX Base Operating System.
- X.25 protocol support requires a separately ordered LPP, IBM AIXLINK/X.25 (5696-926).
- This adapter also functions with AIX 4.2.1 and PSSP 2.4.

9.1.10 ATM 155 TURBOWAYS UTP adapter (F/C 2963)

The TURBOWAYS 155 UTP ATM adapter (F/C 2963) enable TCP/IP applications to work in an asynchronous transfer mode (ATM) environment. This adapter provides dedicated 155 MB per second, full-duplex connection to ATM networks using either Permanent Virtual Circuits (PVC) or ATM Forum compliant Switched Virtual Circuits (SVC) UNI 3.1 signalling. The adapter supports AAL-5 adaptation layer interface and communication with devices located on an ATM network, bridged token ring, Ethernet, or other LAN. LAN Emulation (LANE) is provided by the AIX operating system.

The TURBOWAYS 155 UTP ATM adapter requires customer provided CAT5 High Speed Unshielded Twisted Pair (UTP) or Shielded Twisted Pair (STP) cables. These cables must be certified for ATM operation. Maximum cable length is 100 m, and all cables must be terminated with RJ45 connectors.

Feature characteristics

This feature has the following characteristics:

- 32-bit Bus Master PCI 2.1 adapter
- External RJ45 connector
- Provides signaling channel setup
- Provides virtual connection setup and tear down
- Supports point-to-point and point-to-multipoint switching
- Supports virtual circuits (maximum 1024)
- Supports classical IP and ATRP over ATM (RFC 1577)
- Supports Ethernet LAN Emulation and token ring
- Supports ATM SNMP
- Best effort service

Plugging rules

This feature has the following plugging rules:

- A maximum of two adapters per POWER3 SMP thin node.
- A maximum of four adapters per POWER3 SMP wide node.
- A maximum of two adapters per 332 MHz SMP thin node.
- A maximum of four adapters per 332 MHz SMP wide node.
- This adapter cannot be plugged into slot I5, I6, I7, or I8 in 332 MHz SMP wide nodes.

Attention

The following adapters may be used in any combination, but the total number of these adapters cannot exceed four per SMP wide node: F/C 2741, 2742, 2743, 2963, 2968, and 2988.

Customer components

You must supply the following components with this feature:

- Category 5 High Speed UTP cables (or shielded) with RJ45 connectors (100 m maximum length).
- If you plan to use multipoint connections, you must provide an ATM switch.

Hardware requirement

This feature has the following hardware requirement:

- One PCI adapter slot

Software requirements

This feature requires the following software:

- PSSP 3.1, or later.
- AIX 4.3.2, or later.
- This adapter will also function with PSSP 2.4 and AIX 4.2.1.

9.1.11 Ethernet 10/100 MB (F/C 2968)

The IBM 10/100 Ethernet TX PCI adapter (F/C 2968) is a 10/100 PCI Ethernet adapter that is compatible with IEEE 802.3 and 802.3u specifications. The adapter has one RJ-45 connection that supports connections to 100BaseTX and 10BaseT networks.

Feature characteristics

This feature has the following characteristics and requirements:

- Compatible with IEEE 802.3 Standards
- 32-bit Bus Master PCI Bus 132 MBps
- Supports auto-negotiation of media speed and duplex operation
- Supports both full and half duplex operation over 10BaseT networks via the RJ-45 connector

Feature components

This feature order provides the following:

- Adapter card
- Diskette with adapter device driver
- Installation instructions

Customer components

You must supply the following components for this feature:

- Network equipment, such as a hub or switch, required to attach to 10BaseT Ethernet LANs
- All Ethernet cables

Attention

For 100BaseTX connections, UTP Category 5 cabling is required.

Plugging rules

This feature has the following plugging rules:

- A maximum of two adapters per POWER3 SMP thin node.
- A maximum of four adapters per POWER3 SMP wide node.
- A maximum of two adapters per 332 MHz SMP thin node.
- A maximum of four adapters per 332 MHz SMP wide node.
- The following adapters may be used in any combination, but the total number of these adapters must not exceed four total per SMP wide node: F/C 2741, 2742, 2743, 2963, 2968, and 2988.
- Should not be used in slots I5, I6, I7, or I8 in 332 MHz SMP wide nodes. Performance of this adapter is reduced in those slots.

Hardware requirement

This feature has the following hardware requirement:

- One PCI adapter slot

Software requirements

This feature has the following software requirements:

- AIX 4.2.1, or later, installed on the node
- PSSP 2.4, or later, installed on the node
- Adapter device driver (provided with adapter)

9.1.12 Gigabit Ethernet - SX adapter (F/C 2969)

The PCI Gigabit Ethernet - SX adapter (F/C 2969) is a 1000 Mbps PCI Ethernet adapter that is compatible with IEEE 802.3z specifications. The adapter has one external fiber connection that attaches to 1000BaseSX networks via 50 and 62.5 micron multimode cables with SC connectors.

Feature characteristics

This feature has the following characteristics and requirements:

- Compatible with IEEE 802.3z Standards
- Supports full duplex operation over 1000BaseSX networks
- Supports jumbo frames with AIX 4.3.2 device driver

Feature components

This feature order provides the following:

- Adapter card

- Fiber wrap plug
- Installation instructions

Customer components

You must supply the following components for this feature:

- Network equipment, such as a switch or router, is required to attach to 1000BaseSX networks.
- All Ethernet cables.

Attention

The maximum operating distances for the fiber cables are:

- 260 meters with 62.5 micron multimode fiber
- 440 meters with 50 micron multimode fiber

Plugging rules

Use the following PCI Bus definitions for the Gigabit Ethernet plugging rules:

PCI Bus 1 Slots I2 and I3 (on CPU side)

PCI Bus 2 Slots I1, I2, I3, and I4 (on I/O side)

PCI Bus 3 Slots I5, I6, I7, and I8 (on I/O side)

This feature has the following plugging rules:

- 332 MHz SMP thin nodes
 - A maximum of one adapter per node.
 - If a PCI Gigabit Ethernet adapter is installed, F/C 2741, 2742, 2743, 2963, 2968, and 2988 are not supported.
- 332 MHz SMP wide nodes
 - A maximum of two adapters per node.
 - A maximum of one adapter per PCI bus.
 - The adapters must be placed into PCI Bus 1 and PCI Bus 2-slot I4.
 - The adapter is not supported in PCI Bus 3.
 - If one PCI Gigabit Ethernet adapter is installed in a 332 MHz wide node, a maximum of two F/C 2741, 2742, 2743, 2963, 2968, and 2988 are supported.

Note

These adapters must not be placed on the same PCI Bus with the PCI Gigabit Ethernet adapter.

- If two PCI Gigabit Ethernet adapters are installed, F/C 2741, 2742, 2743, 2963, 2968, and 2988 are not supported.
- POWER3 thin nodes
 - A maximum of one adapter per node.
 - The adapter can exhibit sub-optimal performance when plugged into PCI Bus 1.
 - If a PCI Gigabit Ethernet adapter is installed, F/C 2741, 2742, 2743, 2963, 2968, and 2988 are not supported.
- POWER3 wide nodes
 - A maximum of two adapters per node.
 - A maximum of one adapter per PCI bus.
 - If two adapters are installed, one must be placed in PCI Bus 2, and the other in PCI Bus 3.
 - If one PCI Gigabit Ethernet adapter is installed, a maximum of four F/C 2741, 2742, 2743, 2963, 2968, and 2988 are supported.
Note that the adapters must be distributed over the PCI buses.
 - If two PCI Gigabit Ethernet adapters are installed, a maximum of two F/C 2741, 2742, 2743, 2963, 2968, and 2988 are supported.
Note that the adapters must be distributed over the PCI buses.

Hardware requirements

This feature has the following hardware requirements:

- One PCI 32-bit or 64-bit adapter slot in POWER3 nodes
- One PCI 32-bit adapter slot in 332 MHz nodes

Software requirements

This feature has the following software requirements:

- POWER3 thin and wide nodes
- PSSP 3.1 and AIX 4.3.2 or later
- 332 Mhz thin and wide nodes
- PSSP 3.1 and AIX 4.3.2, or later

9.1.13 Ethernet 10 MB BNC (F/C 2985)

The PCI Ethernet BNC/RJ-45 adapter (F/C 2985) is a 10 Mbps PCI Ethernet adapter that is compatible with IEEE 802.3 specifications. The adapter has two external connections: BNC to attach to 10Base2 networks and RJ-45 to attach to 10BaseT networks.

Feature characteristics

This feature has the following characteristics and requirements:

- 10 Mbps Ethernet compatible with IEEE 802.3 Standards.
- 32-bit Bus Master PCI Bus 132 MBps.
- Supports half duplex operations over 10Base2 networks via the BNC connector
- Supports both full and half duplex operations over 10BaseT networks via the RJ-45 connector

Feature components

This feature order provides the following:

- Adapter card
- RJ-45 and BNC diagnostic wrap plugs
- Installation instructions

Customer components

You must supply the following components for this feature:

- Network equipment, such as a hub or switch, required to attach to 10BaseT Ethernet LANs
- All Ethernet cables

Attention

For 10BaseT connections, UTP Category 3, 4, or 5 cabling is required. UTP Category 5 cabling is strongly recommend to facilitate upgrades to 100 Mbps Ethernet LAN without cabling changes.

Plugging rules

This feature has the following plugging rules:

- A maximum of two adapters per SMP thin node
- A maximum of eight adapters per SMP wide node

Hardware requirement

This feature has the following hardware requirement:

- One PCI adapter slot

Software requirements

This feature has the following software requirements:

- AIX 4.2.1, or later, installed on the node
- PSSP 2.4, or later, installed on the node
- Adapter device driver (part of base AIX BOS code)

9.1.14 Ethernet 10 MB AUI (F/C 2987)

The PCI Ethernet BNC/RJ-45 adapter (F/C 2985) is a 10 Mbps PCI Ethernet adapter that is compatible with IEEE 802.3 specifications. The adapter has two external connections: BNC to attach to 10Base5 networks and RJ-45 to attach to 10BaseT networks.

Feature characteristics

This feature has the following characteristics and requirements:

- 10 MBps Ethernet compatible with IEEE 802.3 Standards
- 32-bit Bus Master PCI Bus 132 MBps
- Supports half duplex operations over 10Base5 networks via the BNC connector
- Supports both full and half duplex operation over 10BaseT networks via the RJ-45 connector

Feature components

This feature order provides the following:

- Adapter card
- RJ-45 and AUI diagnostic wrap plugs
- Installation instructions

Customer components

You must supply the following components for this feature:

- Network equipment, such as a hub or switch, required to attach to 10BaseT Ethernet LANs
- All Ethernet cables

Attention

For 10BaseT connections, UTP Category 3, 4, or 5 cabling is required. UTP Category 5 cabling is strongly recommended to facilitate upgrades to 100 Mbps Ethernet LAN without cabling changes.

Plugging rules

This feature has the following plugging rules:

- A maximum of two adapters per SMP thin node
- A maximum of eight adapters per SMP wide node

Hardware requirements

This feature has the following hardware requirements:

- One PCI adapter slot

Software requirements

This feature has the following software requirements:

- AIX 4.2.1, or later, installed on the node
- PSSP 2.4, or later, installed on the node
- Adapter device driver (part of base AIX BOS code)

9.1.15 ATM 155 MMF (F/C 2988)

The TURBOWAYS 155 ATM adapter (F/C 2988) enables TCP/IP applications to work in an asynchronous transfer mode (ATM) environment. This adapter provides dedicated 155 MB per second, full-duplex connection to ATM networks using either Permanent Virtual Circuits (PVC) or ATM Forum compliant Switched Virtual Circuits (SVC) UNI 3.1 signalling. The adapter supports AAL-5 adaptation layer interface and communication with devices located on an ATM network, bridged token ring, Ethernet, or other LAN. LAN Emulation (LANE) is provided by the AIX operating system.

Feature characteristics and requirements

This feature has the following characteristics and requirements:

- Provides signaling channel setup
- Provides virtual connection set up and tear down
- Supports point-to-point and point-to-multipoint switching
- Supports virtual circuits (maximum 1024)
- Supports classical IP and ATRP over ATM (RFC 1577)

- Supports Ethernet LAN Emulation and Token Ring
- Supports ATM SNMP

Plugging rules

This feature has the following plugging rules:

- A maximum of two adapters per POWER3 SMP thin node.
- A maximum of four adapters per POWER3 SMP wide node.
- A maximum of two adapters per 332 MHz SMP thin node.
- A maximum of four adapters per 332 MHz SMP wide node.
- The following adapters may be used in any combination, but the total number of these adapters must not exceed four total per SMP wide node: F/C 2741, 2742, 2743, 2963, 2968, and 2988.
- The adapter cannot be plugged into slot I5, I6, I7, or I8 in 332 MHz SMP wide nodes.

Customer components

You must supply the following components with this feature:

- Plenum rated 62.5/125 multimode fiber cables terminated with an SC connector
- An ATM switch

Software requirements

This feature requires the following software:

- PSSP 2.4
- AIX 4.2.1, or later

9.1.16 Ultra SCSI Single Ended adapter (F/C 6206)

The PCI SCSI-2 Ultra/Wide Single Ended adapter (F/C 6206) provides a single ended SCSI-2 Ultra/Wide interface that can burst data between devices on the SCSI bus at 40 MBps (twice the fast/wide rate) using block sizes greater than 64 K. It conforms to SCSI-2 standards and Fast-20 (Ultra) documentation. F/C 6206 supports both internal and external devices connected to the same SCSI bus. Industry standard SCSI P (68-pin) connectors are incorporated on the adapter.

Feature characteristics

This feature has the following characteristics:

- 32-bit Bus Master PCI 2.1 adapter

- Supports attachment of internal and external single ended 8-bit and 16-bit SCSI or Ultra SCSI devices
 - External connections on J2 with 68 pin SCSI-3 standard P connector
 - Internal connections on J3 with 68 pin high density SCSI connector for 16-bit attachments
 - Internal connections on J4 with 50 pin (2x25) SCSI connector for 8-bit attachments

Adapter limitations

- Data transfer rates are limited to the speed of the slowest attached device. For example, if you connect an Ultra drive and a fast/wide drive, the adapter will limit data transfers to fast/wide rates.
- If a cable is attached to the external J2 connector, data transfer rates will be limited to fast/wide rates.
- Ultra data transfer rates can only be achieved using the internal connections with cable lengths of 1.5 m or less.
- External cable lengths are limited to 3 m for fast/wide data transfer rates.
- The internal J3 and J4 connectors cannot be used at the same time.

Customer components

You must supply the following components for this feature:

- If you are using F/C 6206 to configure independent internal disk in an 332 MHz SMP wide node, you must also order F/C 1241.

Plugging rules

This feature has the following plugging rules:

- A maximum of two adapters per POWER3 SMP thin node.
- A maximum of four adapters per POWER3 SMP wide node.
 - Distribute evenly across all three slot groups.
- A maximum of two adapters per 332 MHz SMP thin node.
- A maximum of four adapters per 332 MHz SMP wide node.
 - Distribute evenly across the two recommended slot groups.
 - With 332 MHz SMP wide nodes, this adapter cannot be installed in the third PCI slot group (I5, I6, I7, and I8).

Attention

If you are using F/C 6206 along with F/C 1241 to provide an independent SCSI bus (such as for mirroring) in 332 MHz SMP wide nodes, the adapter must be placed either slot I1, I2, I3, or I4 on the I/O side of the node.

Hardware requirement

This feature has the following hardware requirement:

- One PCI adapter slot per adapter

Attention

Single Ended (SE) SCSI adapters cannot interoperate with Differential SCSI adapters in twin-tailed (high availability) configurations.

Software requirements

This feature has the following software requirements:

- AIX 4.3.2, or later, installed on the node.
- PSSP 3.1, or later, installed on the node.
- This adapter also functions with AIX 4.2.1 and PSSP 2.4.

9.1.17 Ultra SCSI Differential (F/C 6207)

The PCI SCSI-2 Ultra/Wide Differential adapter (F/C 6207) provides a differential SCSI-2 Ultra/Wide interface that can burst data between devices on the SCSI bus at 40 MBps. F/C 6207 supports Ultra and Fast/Wide synchronous data transfers, and it supports external devices (no internal connections) up to 25 m away. This adapter conforms to SCSI-2 standards and the Fast-20 (Ultra) documentation. Industry standard SCSI P (68-pin) connectors are incorporated on the adapter.

Attention

Data transfer rates with F/C 6207 are limited to the speed of the slowest device on the SCSI bus.

Feature characteristics

This feature has the following characteristics:

- 32-bit Bus Master adapter.

- Supports attachment of external 8-bit or 16-bit SCSI devices on the J2 port using a 68 pin SCSI-3 standard connector.

Customer components

Optional cables are available through IBM.

Plugging rules

This feature has the following plugging rules:

- A maximum of two adapters per POWER3 SMP thin node.
- A maximum of four adapters per POWER3 SMP wide node.
 - Distribute evenly across all three slot groups.
- A maximum of two adapters per 332 MHz SMP thin node.
- A maximum of four adapters per 332 MHz SMP wide node.
 - Distribute evenly across the two recommended slot groups.
 - With 332 MHz SMP wide nodes, this adapter cannot be installed in the third PCI slot group (I5, I6, I7, and I8).

Hardware requirement

This feature has the following hardware requirement:

- One PCI adapter slot per adapter

Attention

Single Ended (SE) and Double Ended SCSI adapters cannot be twin-tailed to the same external disk array when used in a high-availability configuration.

Software requirements

This feature has the following software requirements:

- AIX 4.3.1 or later installed on the node.
- PSSP 3.1 or later installed on the node.
- This adapter also functions with AIX 4.2.1 and PSSP 2.4.

9.1.18 SCSI-2 F/W Single-Ended adapter (F/C 6208)

The PCI SCSI-2 Fast/Wide Single Ended adapter (F/C 6208) provides a single ended SCSI-2 Fast/Wide interface that can burst data between devices on the SCSI bus at 20 MBps. It conforms to SCSI-2 standards and supports Fast/Wide synchronous data rates of up to 10 MHz. F/C 6208 supports both internal and external devices connected to the same SCSI bus.

Feature characteristics

This feature has the following characteristics:

- 32-bit Bus Master adapter
- Supports attachment of internal and external single ended 8-bit and 16-bit SCSI devices

External connections on J2 with 68 pin SCSI-3 standard P connector

Internal connections on J3 with 68 pin high density SCSI connector for 16-bit attachments and on J4 with 50 pin SCSI connector for 8-bit attachments

Attention

The J3 and J4 connectors cannot be used at the same time.

Customer components

You must supply the following components for this feature:

- If you are using F/C 6208 to connect internal DASD in a 332 MHz SMP wide node, you must also order F/C 1241.

Plugging rules

This feature has the following plugging rules:

- A maximum of two adapters per SMP thin node
- A maximum of eight adapters per SMP wide node

If you are using F/C 6208 for mirroring (F/C 1241) in 332 MHz SMP wide nodes, the adapter must be placed either slot I1, I2, I3, or I4 on the I/O side of the node.

Hardware requirement

This feature has the following hardware requirement:

- One PCI adapter slot per adapter

Attention

Single Ended (SE) SCSI adapters cannot interoperate with Differential SCSI adapters in twin-tailed (high availability) configurations.

Software requirements

This feature has the following software requirements:

- AIX 4.2.1, or later, installed on the node

- PSSP 2.4, or later, installed on the node

9.1.19 SCSI-2 F/W Differential adapter (F/C 6209)

The PCI SCSI-2 Fast/Wide Differential adapter (F/C 6209) provides a differential SCSI-2 Fast/Wide interface that can burst data between devices on the SCSI bus at 20 MBps. It conforms to SCSI-2 standards and supports Fast/Wide synchronous data rates of up to 10 MHz. F/C 6209 supports external devices connected to the same SCSI bus.

Feature characteristics

This feature has the following characteristics:

- 32-bit Bus Master adapter
- Supports attachment of external 16-bit SCSI devices on the J2 port using a 68 pin SCSI-3 standard P connector

Customer components

None.

Plugging rules

This feature has the following plugging rules:

- A maximum of two adapters per SMP thin node
- A maximum of eight adapters per SMP wide node

Hardware requirement

This feature has the following hardware requirement:

- One PCI adapter slot per adapter

Attention

Single Ended (SE) and Double Ended SCSI adapters cannot be twin-tailed to the same external disk array when used in a high-availability configuration.

Software requirements

This feature has the following software requirements:

- AIX 4.2.1, or later, installed on the node
- PSSP 2.4, or later, installed on the node

9.1.20 SSA RAID 5 adapter (F/C 6215)

The PCI SSA RAID 5 adapter supports RAID 5 SSA disk arrays and can be used to access non-RAID disks between multiple hosts. It has the capability to improve write response time in the single initiator mode for both RAID and non-RAID disks by the addition of the Fast-Write Cache Option (F/C 6222). For more details on the F/C 6222 option, refer to 9.1.21, “SSA Fast-Write Cache Module (F/C 6222)” on page 166.

Feature characteristics and requirements

This feature has the following characteristics and requirements:

- 32-bit PCI bus.
- Support for floating hot spares on the same loop.
- RAID 5 arrays from (2+P) up to (15+P).
- Up to 6 (15+P) or 32 (2+P) RAID 5 Array Groups per adapter.
- All members of a RAID 5 array must be on the same SSA loop.

Plugging rules

This feature has the following plugging rules:

- A maximum of two adapters per SMP thin node.
- A maximum of six adapters per SMP wide node.
- The adapter cannot be plugged into slot I5, I6, I7, or I8 in 332 MHz SMP wide nodes.

Attention

When more than one SSA RAID adapter is used, the adapters should be distributed evenly across all available PCI slots.

Software requirements

This feature has the following software requirements:

- AIX 4.2.1 or later installed on the node
- PSSP 2.4 or later installed on the node
- Adapter device driver and FDDI common code (provided with adapter)

9.1.21 SSA Fast-Write Cache Module (F/C 6222)

The SSA Fast-Write Cache is an optional 4 MB fast-write module that plugs into the PCI SSA RAID 5 adapter (F/C 6215). The F/C 6222 cache option uses non-volatile RAM having over seven years of memory retention.

Non-volatile memory allows you to transfer the cache module from a failing Multi-Initiator adapter to a new adapter during the unlikely event of an adapter failure. This helps insure data integrity and operational reliability.

Feature characteristics and requirements

This feature has the following characteristics and requirements:

- Only one F/C 6222 is supported on each PCI SSA RAID 5 adapter (F/C 6215).
- Requires PSSP 2.4 or greater and either AIX 4.2.1 or later.

9.1.22 ARTIC960RxD Quad Digital Trunk adapter (F/C 6310)

The ARTIC960RxD Quad Digital Trunk adapter provides voice processing for up to four T1 or E1 digital trunk lines, providing connectivity for 96 (T1) or 120 (E1) voice channels in a single PCI slot. The voice processing function is provided by DirectTalk for AIX, Version 2.1 LPP. The adapter provides high-function control of I/O operations and serves to off-load I/O tasks from the system microprocessor.

Feature characteristics

This feature has the following characteristics:

- 32-bit PCI 2.1 adapter
- One 36-pin, high-density port
- Support for up to four T1 or E1 trunk lines
- Supports voice processing using DirectTalk for AIX

Plugging rules

This feature has the following plugging rules:

- POWER3 thin nodes
 - A maximum of two adapters per node.
- POWER3 wide nodes
 - A maximum of four adapters per node.
- 332 MHz thin nodes
 - A maximum of two adapters per node.
- 332 MHz wide nodes
 - A maximum of four adapters per node.
 - This adapter is not supported in PCI bus slots I5, I6, I7, or I8 on the I/O side of the 332 Mhz wide node.

Feature components

This feature order provides the following:

- One ARTIC960RxD adapter (F/C 6310).
- A connecting cable (required). The following are available from IBM:
 - F/C 2709** ARTIC960Hx 4-port T1 RJ45 cable
 - F/C 2710** ARTIC960Hx 4-port E1 RJ45 cable
 - F/C 2871** ARTIC960RxD Quad DTA, T1, 100 ohm, 3 m 4-port cable
 - F/C 2872** ARTIC960RxD Quad DTA, T1, 100 ohm, 15 m extension cable
 - F/C 2873** ARTIC960RxD Quad DTA, E1, 120 ohm balanced, 3 m 4-port cable
 - F/C 2874** ARTIC960RxD Quad DTA, E1, 120 ohm balanced, 7.5 m extension cable
 - F/C 2875** ARTIC960RxD Quad DTA, E1, 75 ohm unbalanced-grounded, 1.8 m 4-port cable
 - F/C 2876** ARTIC960RxD Quad DTA, E1, 75 ohm unbalanced-ungrounded, 1.8 m 4-port cable
 - F/C 2877** ARTIC960RxD Quad DTA, H.100, 4-drop cable

Hardware requirement

This feature has the following hardware requirement:

- One 32-bit PCI adapter slot

Software requirements

This feature has the following software requirements:

- AIX 4.2.1, AIX 4.3.2, or later
- DirectTalk for AIX, Version 2.1 LPP (5765-B81) to provide voice processing
- Adapter device driver (provided with adapter)

9.1.23 IBM Network Terminal Accelerator (256 Session - F/C 2402)

The IBM Network Terminal Accelerator feature (F/C 2402) is an Ethernet adapter that accelerates network performance by off-loading the telnet and rlogin daemons, TCP/IP protocol stack and virtual terminal I/O management from the RS/6000 system. The network adapter buffers the system from frequent CPU intensive packet interrupts, increases terminal I/O throughput and the number of concurrent online user sessions, and reduces context switches, which dramatically reduces the CPU load.

The network adapter software provides a pass-through capability for other Ethernet protocols, which can eliminate the need for a separate Ethernet adapter. The network adapter supports onboard simple network management protocol (SNMP) for network management.

9.1.24 IBM Network Terminal Accelerator (2048 Session - F/C 2403)

The IBM Network Terminal Accelerator feature (F/C 2403) is an Ethernet adapter that accelerates network performance by off-loading the telnet and rlogin daemons, TCP/IP protocol stack, and virtual terminal I/O management from the RS/6000 system. The adapter buffers the system from frequent CPU intensive packet interrupts, increases terminal I/O throughput and the number of concurrent online user sessions by up to three times, and reduces context switches, which dramatically reduces the CPU load.

The network adapter software provides a pass-through capability for other Ethernet protocols, which can eliminate the need for a separate Ethernet adapter. The network adapter supports onboard SNMP for network management.

9.1.25 SCSI-2 High Performance External I/O Controller (F/C 2410)

The SCSI-2 External I/O Controller feature (F/C 2410) allows you to attach external single-ended SCSI and SCSI-2 devices. This feature provides for attachment of one IBM 9334 Expansion Unit Model 500 or up to four external IBM supported SCSI devices with IBM supported cables.

9.1.26 Enhanced SCSI-2 Differential Fast/Wide adapter/A (F/C 2412)

The IBM Enhanced SCSI-2 Differential Fast/Wide adapter/A is a dual ported fast (10 MHz) and wide (2 bytes wide) SCSI Micro Channel adapter that can provide synchronous SCSI bus data rates of up to 20 megabytes per second.

This adapter provides high performance attachment to Differential SCSI disks, disk subsystems, tape devices, and read/write optical subsystems. The maximum data rate depends on system and application configurations. This adapter has one internal single ended port and one external differential port. The internal port is capable of attaching up to six single ended devices. The external port is capable of addressing up to fifteen differential devices. The number of physical devices attached to each port is limited by SCSI bus cabling restrictions. The internal port of this adapter supports either 8-bit or 16-bit devices via an 8-bit or a 16-bit connector. Only one of these two connectors may be used at one time. (Devices of different bus attachment widths cannot be connected/used at the same time.) The external Differential SCSI bus is capable of supporting cable lengths of 25 meters (82 feet).

Additional system, subsystem, and high availability connections are also available with the differential system-to-system and Y-cable features.

9.1.27 SCSI-2 Fast/Wide adapter/A (F/C 2415)

The SCSI-2 Fast/Wide adapter feature (F/C 2415) is a dual-ported fast (10 MHz) and wide (two bytes) adapter. It provides synchronous SCSI bus rates up to 20 megabytes per second and attaches to single-ended (SE) SCSI disks, CD-ROMs, tape drives, R/W optical devices, and storage subsystems. The maximum data rate depends on the maximum rate of the attached device.

This adapter has one internal SE port and one external SE port. Each SE port can address up to seven SE SCSI devices. The number of physical devices attached to each port is limited by SCSI bus cabling restrictions. The internal port supports either 8-bit or 16-bit devices via an internal fast/wide cable with an interposer for fast-only devices. External cabling may be up to six meters (19.6 feet) when attached to the 9334-010 or 9334-500, or three meters when attached to anything else, and is supplied by the attaching device.

9.1.28 SCSI-2 Differential Fast/Wide adapter/A (F/C 2416)

The SCSI-2 Differential Fast/Wide feature (F/C 2416) is a dual-ported fast (10 MHz) and wide (two bytes) adapter. It provides synchronous SCSI bus rates up to 20 Mbps and attaches to SCSI fixed disks, CD-ROM devices, tape drives, R/W optical devices, and storage subsystems. The maximum data rate depends on the maximum rate of the attached device.

9.1.29 SCSI-2 Differential External I/O Controller (F/C 2420)

The SCSI-2 Differential External I/O Controller feature (F/C 2420) allows you to attach external SCSI-2 differential devices. This adapter provides SCSI bus signal cable quality and a maximum SCSI bus length of up to 19 meters (62.3 feet).

9.1.30 4-Port Multiprotocol Communications Controller (F/C 2700)

The 4-port Multiprotocol Communications Controller feature (F/C 2700) attaches the RS/6000 500 series to synchronous communications networks using EIA-232D, EIA-422A, A.35, and X.21 physical specifications. The adapter supports SDLC and BSC protocols, prepares all inbound and outbound data, performs address searches, and in general, relieves the system processor of many communications tasks. It is designed to support data rates up to 64 Kbps per port with appropriate user provided software.

9.1.31 FDDI attachment (F/C 2724 and F/C 2723)

The SP supports both single-ring (F/C 2724) and dual-ring (F/C 2723) attachments.

The FDDI single-ring attachment station (SAS) adapter attaches the SP directly to a primary ring of an FDDI network through a concentrator. The FDDI concentrator offers additional protection by isolating the network from routine on/off activity and individual failure of an SP processor node.

9.1.32 HIPPI (F/C 2735)

The High Performance Parallel Interface (HIPPI) (F/C 2735) provides high-speed connectivity to super computers, RS/6000 processors, HIPPI fiber optic extenders, IBM 9570 Disk Array, and other disk arrays and tape systems.

This feature provides an efficient simplex/duplex point-to-point HIPPI interface achieving peak rates of 100 megabytes per second (simultaneous in each direction) over a distance of up to 25 meters via copper cabling. This distance can be extended using HIPPI extenders. The adapter can be used for either communication or storage-channel applications.

The HIPPI adapter set occupies three adjacent Micro Channel slots. However, because of power considerations, the adapter set currently must be considered to occupy five Micro Channel slots.

9.1.33 S/390 ESCON Channel Emulator adapter (F/C 2754)

This adapter provides attachment capability via the IBM ESCON architecture for selected tapes providing IBM customers with more choices for implementing data access applications by an ESCON Channel attachment of S/390 tapes to RS/6000 systems. Supporting a data transfer rate of up to 17 MB per second (Mbps), the ESCON Emulator adapter allows attachment of ESCON attached tape subsystems. The adapter uses two Micro Channel slots. A maximum of two adapters may be installed per processor, depending upon slot availability. Designed to support specifications for ESCON devices, the ESCON Emulator adapter conforms to most of the standard Micro Channel specifications that are required for tape subsystems. One wrap plug, two diagnostic diskettes (stand-alone and runtime), publications, and two device driver diskettes are included with the hardware adapter. Channel cables are also required and should be ordered separately.

9.1.34 Block Multiplexer Channel adapter - BMCA (F/C 2755)

A DB78 bus/tag terminator is shipped with F/C 2753 to end the bus and tag channel string so that you do not need to supply serpentine bus and tag terminators for the channels connected to the SP BMCA feature (F/C 2755).

9.1.35 ESCON Control Unit adapter (F/C 2756)

This adapter (F/C 2756) allows you the ability to attach SP nodes to the IBM Enterprise System Connection (ESCON) channels of the System/390. The adapter attaches directly to an ESCON channel providing fiber optical links using LED technology. It also attaches to ESCON Directors (fiber optic switches) to allow for large numbers of connections.

9.1.36 8-Port Async adapter - EIA-232 (F/C 2930)

The 8-port Async feature (F/C 2930) provides the RS/6000 500 series system with up to eight EIA-232 asynchronous serial devices such as terminals and printers. The 8-port Async adapter contains all of the electronics required to support eight asynchronous ports and uses one I/O card slot.

9.1.37 8-Port Async adapter - EIA-422A (F/C 2940)

The 8-port Async feature (F/C 2940) provides the RS/6000 500 series system with up to eight EIA-422A asynchronous serial devices such as terminals and printers. The 8-port Async adapter contains all of the electronics required to support eight asynchronous ports and uses one I/O card slot.

9.1.38 X.25 Interface Co-Processor/2 (F/C 2960)

The X.25 Interface Co-Processor/2 feature (F/C 2960) attaches the RS/6000 500 series to an X.25 Packet Switched network. The X.25 adapter provides a single port that accommodates one of the following selectable interfaces: X.21, EIA-232D/V.24, and V.35. This adapter allows the systems to be attached to an X.25 network, and its on-board software is capable of processing inbound and outbound data streams to offload communications tasks from the system processor.

9.1.39 Token Ring High Performance Network adapter (F/C 2970)

The Token Ring High Performance Network adapter (F/C 2970) is designed to allow an SP node to attach to 4 Mbps or 16 Mbps Token Ring local area network. This adapter is cable-and-network compatible with all IBM PS/2 Token Ring adapters. The required cable is included with the adapter and is 20 feet in length. Extension cables may be ordered separately.

9.1.40 Auto Token-Ring LANstreamer MC 32 adapter (F/C 2972)

The IBM Auto Token Ring LANstreamer MC 32 feature (F/C 2972) is designed to allow an RS/6000 system to attach to 4 Mbps or 16 Mbps token ring local area networks. The adapter automatically selects the correct token ring speed (4 or 16 Mbps). It is cable and network compatible with all IBM PS/2 Token ring adapters, which means that no new cables or network components are required.

9.1.41 Ethernet High Performance LAN adapter (F/C 2980)

The Ethernet High Performance LAN adapter (F/C 2980) is a high performance MCA architecture Busmaster adapter that provides a connection to 10 MB Carrier Sense Multiple Access/Collision Detection (CSMA/CD) Ethernet networks. The primary use of this adapter is to attach the 9076 system to Ethernet networks. F/C 2980 has both a 10Base2 (BNC) connector and a 10Base5 (15 pin, thick) connector, but only one connector may be used at one time.

9.1.42 TURBOWAYS 100 ATM adapter (F/C 2984)

The TURBOWAYS 100 ATM adapter (F/C 2984) enables TCP/IP applications to work in an asynchronous transfer mode (ATM) environment. One virtual connection is dedicated to each IP address, and a transformation of each IP address to the corresponding virtual connection is performed.

The initial release supports AAL-5 adaptation layer interface and supports 1024 active virtual connections.

9.1.43 TURBOWAYS 155 ATM adapter (F/C 2989)

The TURBOWAYS 155 ATM adapter (F/C 2989) enables TCP/IP applications to work in an asynchronous transfer mode (ATM) environment. One virtual connection is dedicated to each IP address, and a transformation of each IP address to the corresponding virtual connection is performed.

The initial release supports AAL-5 adaptation layer interface and supports 1024 active virtual connections.

9.1.44 Ethernet LAN adapter (AUI/10BaseT) (F/C 2992)

This adapter allows the RS/6000 SP system to attach to 10 Mbps Ethernet networks. F/C 2992 provides both an AUI port and a 10BaseT (RJ-45) Ethernet connection. Only one of the two ports may be used at one time. This adapter has a parallel processing design, which reduces latency and increases data throughput.

9.1.45 Ethernet LAN adapter 10Base2 (BNC) (F/C 2993)

This adapter allows the RS/6000 SP system to attach to 10 Mbps Ethernet networks. F/C 2993 provides a 10Base2 (BNC) Ethernet connection. This adapter has a parallel processing design, which reduces latency and increases data throughput.

9.1.46 10/100 Ethernet Twisted Pair MC adapter (F/C 2994)

The 10/100 Ethernet twisted pair MC adapter allows the RS/6000 SP system to attach to both 100Base-TX (IEEE 802.3u) and 10Base-T (IEEE 802.3) Ethernet networks. The adapter automatically senses network transfer rates and selects the appropriate rate at power-up. F/C 2994 provides network attachment through a single, RJ-45 port that supports category 5 unshielded twisted pair (UTP) wiring for 100Base-TX connections and category 3, 4, or 5 UTP wiring for 10Base-T connections. Type 100 VG wiring is not supported.

If your network currently operates at 10 Mbps, and your plans include migration to 100 Mbps operation, you should consider using a category 5 cable now.

9.1.47 Ethernet 10BaseT Transceiver (F/C 4224)

The Ethernet 10BaseT Transceiver feature (F/C 4224) provides the complete attachment unit interface (AUI) to a twisted pair LAN connection.

9.1.48 9333 High Performance Subsystem adapter (F/C 6212)

The 9333 High Performance Subsystem adapter (F/C 6212) allows attachment of four (per adapter) 9333 High Performance Disk Drive Subsystems to an RS/6000 processor.

9.1.49 SSA 4-Port adapter (F/C 6214)

The SSA 4-Port adapter (F/C 6214) provides Serial Storage Architecture (SSA) connections that can be configured to provide two SSA loops. Each loop will support the attachment of up to 48 devices (96 devices per adapter). Each adapter will support attachment of up to six maximum configuration IBM 7133 Serial Storage Architecture Disk Subsystems (96 drives) for a total disk drive capacity of 432 GB using 4.5 GB disk drives.

9.1.50 Enhanced SSA 4-Port adapter (F/C 6216)

The Enhanced SSA 4-Port adapter serves as an interface between systems using Micro Channel Architecture (MCA) and devices using Serial Storage Architecture (SSA). F/C 6216 provides 4 SSA ports for the attachment of data

storage devices. The adapter's four ports are arranged in two configurable pairs providing two SSA loops. Each loop will support the attachment of 48 devices or 96 devices per adapter card. F/C 6216 also supports six IBM 7133 SSA Subsystems per adapter. This permits attaching up to 96 disk drives for a storage capacity of 432 GB per adapter when using 4.5 GB disk drives.

9.1.51 SSA 4-Port RAID adapter (F/C 6217)

The SSA 4-Port RAID adapter (F/C 6217) is a new addition to the SSA family of adapters for SP systems. The SSA 4-Port adapter offers the Redundant Array of Independent Disks (RAID) 5 function, which provides protection to your data in the event of a disk drive failure. This adapter also supports attachment to a non-RAID disk in a single-initiator per loop environment. A utility program is provided to control the RAID configuration.

9.1.52 SSA Multi-Initiator/RAID EL adapter (F/C 6219)

The Micro Channel SSA Multi-Initiator/RAID EL adapter can be configured as either a two initiator non-RAID adapter or as a one initiator RAID 5 adapter. This adapter has four ports and two SSA loops supporting 48 SSA disk drives per loop (96 drives per adapter). F/C 6219 also supports HACMP functions.

9.1.53 SSA Fast-Write Cache Module (F/C 6222)

The Micro Channel SSA multi-initiator/RAID EL adapter (F/C 6219) supports a 4 MB Fast-Write Cache option (F/C 6222) that improves write performance in both the RAID 5 and non-RAID configurations.

9.2 SP-attached servers communication adapters

This section contains information about supported communication adapters for SP-attached servers.

Table 25 shows a list of supported communication adapters. Refer to 9.1, "PCI nodes communication adapters" on page 137, for a detailed description of each adapter.

Table 25. Supported communication adapters for SP-attached servers

F/C	PCI Adapter Name
2741	FDDI SK-NET LP SAS
2742	FDDI SK-NET LP DAS
2743	FDDI SK-NET UP SAS

F/C	PCI Adapter Name
2751	S/390 ESCON Channel adapter
2920	Token Ring Auto Lanstream
2943	EIA 232/RS-422 8-port Asynchronous adapter
2944	WAN RS232 128-port
2947	IBM ARTIC960Hx 4-Port Selectable adapter
2962	2-port Multiprotocol X.25 adapter
2963	ATM 155 TURBOWAYS UTP adapter
2968	Ethernet 10/100 MB
2969	Gigabit Ethernet - SX
2985	Ethernet 10 MB BNC
2987	Ethernet 10 MB AUI
2988	ATM 155 MMF
6206	Ultra SCSI Single Ended
6207	Ultra SCSI Differential
6208	SCSI-2 F/W Single-Ended
6209	SCSI-2 F/W Differential
6215	SSA RAID 5 (accepts optional SSA Fast-Write Cache module (F/C 6222))
6310	IBM ARTIC960RxD Quad Digital Trunk adapter

9.3 SP Switch Routers network media cards

This section contains information for the SP Switch Router network media cards. These network media cards are used to connect your SP system to an external network using the SP Switch Router.

The features listed in Table 26 are described briefly in the following sections.

Table 26. SP Switch Router network media cards

F/C	Description
1101	ATM OC3, two port SM fiber
1102	ATM OC3, two port MM fiber

F/C	Description
1103	SONET/IP OC3, one port MM fiber
1104	SONET/IP OC3, one port SM fiber
1105	ATM OC12, one port SM fiber
1106	FDDI, four port MM fiber
1107	Ethernet 10/100Base-T, eight port
1108	HIPPI, one port
1109	HSSI, two port
1112	Ethernet 10/100Base-T, four port
1113	Blank faceplate
1114	64 MB DRAM SIMM
1115	ATM OC12, one port MM fiber
4021	SP Switch Router adapter ¹
9310	SP Switch Router adapter cable, 10 meter option (includes 10 m frame-to-frame ground cable)
9320	SP Switch Router adapter cable, 20 meter option (includes 20 m frame-to-frame ground cable)
¹ Choice of either F/C 9310 or F/C 9320 is included with each F/C 4021.	

9.3.1 ATM OC3, two port SM fiber (F/C 1101)

The ATM OC-3c IP Forwarding Media Card is an intelligent, self-contained IP forwarding engine that provides high-performance ATM OC-3c support for the SP Switch Router.

9.3.2 ATM OC3, two port MM fiber (F/C 1102)

The ATM OC-3c IP Forwarding Media Card is an intelligent, self-contained IP forwarding engine that provides high-performance ATM OC-3c support for the SP Switch Router.

9.3.3 SONET/IP OC3, one port MM fiber (F/C 1103)

The SONET OC-3c IP Forwarding Media Card is an intelligent, self-contained IP forwarding engine that provides high-performance SONET OC-3c support for the SP Switch Router.

9.3.4 SONET/IP OC3, one port SM fiber (F/C 1104)

The SONET OC-3c IP Forwarding Media Card is an intelligent, self-contained IP forwarding engine that provides high-performance SONET OC-3c support for the SP Switch Router.

9.3.5 ATM OC12, one port SM fiber (F/C 1105)

The ATM OC-12c IP Forwarding Media Card is an intelligent, self-contained IP forwarding engine that provides high-performance ATM OC-12c support for the SP Switch Router.

9.3.6 FDDI, four port MM fiber (F/C 1106)

The FDDI IP Forwarding Media Card is an intelligent, self-contained IP forwarding engine that provides high-performance FDDI support for the SP Switch Router.

9.3.7 Ethernet 10/100Base-T, eight port (F/C 1107)

The 10/100Base-T IP Forwarding Media Card is an intelligent, self-contained IP forwarding engine that provides high-performance Ethernet support for the SP Switch Router.

9.3.8 HIPPI, one port (F/C 1108)

The HIPPI IP Forwarding Media Card is an intelligent, self-contained IP forwarding engine that provides high-performance HIPPI support for the SP Switch Router.

9.3.9 HSSI, two port (F/C 1109)

Media Card is an intelligent, self-contained IP forwarding engine that provides high-performance HSSI support for the SP Switch Router.

9.3.10 Ethernet 10/100Base-T, four port (F/C 1112)

The 10/100Base-T IP Forwarding Media Card is an intelligent, self-contained IP forwarding engine that provides high-performance Ethernet support for the SP Switch Router.

9.3.11 Blank faceplate (F/C 1113)

F/C 113 provides a blank faceplate that may be needed if a media card is not required or if one is removed from an SP Switch Router.

9.3.12 64 MB DRAM SIMM (F/C 1114)

F/C 1114 provides memory in 64MB increments (2 x 32 MB DRAM) to a maximum of 256 MB per SP Switch Router.

9.3.13 ATM OC12, one port MM fiber (F/C 1115)

The ATM OC-12c IP Forwarding Media Card is an intelligent, self-contained IP forwarding engine that provides high-performance ATM OC-12c support for the SP Switch Router.

9.3.14 SP Switch Router adapter (F/C 4021)

The SP Switch Router adapter is used to connect the SP Switch to an SP Switch Router. You must have one SP Switch Router adapter for each SP system or SP system partition you are going to connect to the SP Switch Router. Each F/C 4021 includes your choice of either F/C 9310 or 9320.

9.3.15 SP Switch Router adapter cable - 10 Meter (F/C 9310)

IBM provides a 10 meter cable for connecting each SP Switch Router adapter to an SP Switch. Each F/C 9310 includes a 10 m frame-to-frame ground cable.

9.3.16 SP Switch Router adapter cable - 20 Meter (F/C 9320)

This optional 20 meter cable is available for connecting each SP Switch Router adapter to an SP Switch. Each F/C 9320 includes a 20 m frame-to-frame ground cable.

Chapter 10. Software support

This chapter describes the software available for RS/6000 SP systems. It covers the following software:

- Parallel System Support Programs (PSSP)
- General Parallel File System (GPFS)
- LoadLeveler
- Parallel Engineering and Scientific Subroutine Library (ESSL)
- Parallel Optimization Subroutine Library (OSL)
- Parallel Environment (PE)
- Performance Toolbox Parallel Extensions (PTPE)

10.1 Parallel System Support Programs (PSSP)

This section describes Parallel System Support Programs (PSSP).

10.1.1 Advantages

The following are advantages of the PSSP:

- A full suite of system management applications with the unique functions required to manage the RS/6000 SP system.
- Simplified installation, operation, and maintenance of all nodes in an RS/6000 SP system. You can operate from a single control workstation.
- Advanced error detection and recovery features that reduce the impact and occurrence of unplanned outages.
- Parallel system management tools for allocating SP resources across the enterprise provided.
- Coexistence is allowed for several releases within an SP partition allowing for easier software migration.
- Advanced performance monitoring for consolidated analysis and reporting.

10.1.2 Description

Parallel System Support Programs (PSSP) is a collection of administrative and operational software applications that run on each node of an RS/6000 SP as well as on SP-attached servers. Built upon the system management tools and commands of the AIX Version 4 operating system, PSSP enables

system administrators and operators to better manage SP systems and their environments.

Sets of software tools and related utilities, including application programming interfaces (APIs), have been grouped together to offer easier administration of installation, configuration, device management, security administration, error logging, system recovery, and resource accounting in the SP environment.

Single point of control

PSSP allows the system administrator/operator to perform all local and remote administrative functions from the control workstation.

Administration and operation

The PSSP system administration and operation component packages together the tools required for SP administrative functions. These include entering and changing configuration information, such as:

- System Data Repository (SDR) for storing management data that can be retrieved across the control workstation, file servers, and SP nodes.
- System command execution for executing system commands in parallel.
- Parallel system management tools and commands for enabling concurrent parallel performance of system management functions across multiple SP nodes.
- File collections for managing duplicated files and directories on multiple nodes.
- Login control for blocking unauthorized user or group access to a specific SP node or a set of nodes.
- Consolidated accounting for centralizing records at the node level (for tracking use by wall clock time rather than processor time) and gathering statistics on parallel jobs. It enhances AIX node data accounting and consolidates information with summaries by node classes. Provides accounting hooks to charge for exclusive node use.

System availability

To significantly improve system availability, PSSP also contains functions and interfaces to other products that can help reduce unplanned outages and minimize the impact of outages that do occur. These include:

- System partitioning, which makes it possible to create a separate logical system partition. It allows testing of different levels and PTFs on the operating system, system software, or hardware. It also allows different

production environments to be concurrently executed for workload isolation.

- Installation and migration coexistence that can reduce scheduled maintenance time. The installation process has been restructured to improve verification between installation steps. Administrators can now more effectively utilize the Network Installation Manager (NIM) functions in AIX. PSSP can coexist with several releases within an SP partition, thus, allowing easier migration to new software levels.
- Node isolation, which removes an SP node from active duty and enables it to be reintegrated without causing an SP Switch fault or disrupting switch traffic. This isolation is useful for correcting an error condition or installing new hardware and software without impacting production.
- High Availability Control Workstation (HACWS) connects two RS/6000 workstations (with HACMP installed) to an SP system to provide a backup control workstation in the event the primary one becomes unavailable (only one is active at any time). A twin-tailed disk configuration, along with the IP address takeover afforded by HACMP, enables rapid switch over to the backup control workstation with little or no impact on operational access to the SP system or System Data Repository.

System monitoring and control

PSSP provides system management tools that enable the system administrator to manage the RS/6000 SP system. PSSP provides the following:

- Enables the system administrator to gracefully shutdown rather than reboot nodes or the complete system
- Allows authorized users to monitor and manipulate SP hardware variables at node or frame levels
- Provides for consolidation of error and status logs for expedited problem determination

New in PSSP 3.2

All PSSP 3.2 components except the PTPE have been enhanced to use the security services. SP security services new capabilities include:

- Kerberos V4 is now optional and is not mandatory to install, configure, or use.
- PSSP supplies commands to configure PSSP use of the DCE security services on the control workstation. For the nodes, PSSP install code installs and configures DCE for you.

- DCE authentication can be used for AIX remote commands that requires root.
- DCE can also be used for trusted services authentication and for AIX remote command authentication.

RS/6000 cluster technology

The RS/6000 Cluster Technology is a collection of services that define hardware and software resources, node relationships, and coordinated actions to manage groups of SP nodes. It provides improved infrastructures for event monitoring and recovery coordination. API for building highly available distributed applications is also provided.

Topology Services defines the relationships between nodes in a cluster in order to allow seamless takeover of functions in the event of a node failure. Group Services provides a set of interfaces that enable a distributed subsystem, such as HACMP/ES or General Parallel File System (GPFS), to synchronize recovery actions among the processes making up the subsystem. Event Management monitors hardware and software resources in the SP and notifies an interested application or subsystem when the state of the resource changes. Resources on any node in the SP can be monitored from any other node.

SP Perspectives

SP Perspectives, a consolidated system graphical user interface, provides a common launch pad for PSSP system management applications through direct manipulation of system objects represented as icons.

This interface is tightly integrated with the problem management infrastructure. It allows users to easily create and monitor system events and provide notification when events occur.

The interface is highly scalable for large systems and can be easily customized to accommodate varying environments.

Offering convenient services

To help ease administrative burdens, PSSP also includes the following publicly available software packages:

- Perl programming language for developing system-wide shell scripts
- Software Update Protocol (SUP) for distributing files from the boot file server
- Kerberos Version 4 security for authentication of the execution of remote commands

- Tool command language (Tcl) for controlling and extending applications

Network Time Protocol (NTP) replacement

There are several ways you might currently be handling synchronization time-of-day clocks on your control workstation and processor nodes. You might already be using Network Time Protocol (NTP) either locally or through the Internet, or you might be using some other time service software.

In PSSP 3.2, the public domain NTP 3.3 has been replaced with AIX NTP 3.4. The rationale behind the replacement of NTP 3.3 is that it is not NLS-enabled. AIX NTP 3.4 is NLS-enabled.

In order to support AIX NTP 3.4 in PSSP 3.2, some changes have been made:

- AIX NTP 3.4 is supported under SRC (System Resource Controller). Internal changes made to start/stop *xntpd* using *src* commands.
- No longer ship public domain NTP 3.3 with PSSP filesets.
- PSSP NTP support configured using *spsiteenv* or *smit site_env_dialog* (no changes required).

For more information on how to configure NTP, refer to IBM Parallel System Support Programs for AIX: Administration Guide, SA22-7348

Subsystem communications support

The Communication Subsystems Support component contains SP Switch adapter diagnostics, switch initialization and fault-handling software, device driver and configuration methods (*config/unconfig*), plus parallel communications APIs.

IBM Virtual Shared Disk

The IBM Virtual Shared Disk (IBM VSD) allows multiple nodes to access a read disk as if the disk were attached locally to each node.

The IBM VSD component is an API that creates logical disk volumes for parallel application access of a real disk device. These can be attached locally or on another SP node. This feature can enhance the performance of applications that provide concurrent control for data integrity, such as Oracle databases.

IBM Recoverable Virtual Shared Disk

The IBM Recoverable Virtual Shared Disk (IBM RVSD) allows transparent recovery of IBM VSD.

The IBM RVSD function provides recovery from failures of IBM VSD server nodes and takes advantage of the availability services provided by PSSP to determine which nodes are up and operational.

Collecting performance data

The Performance Toolbox Parallel Extensions (PTPE) collects and displays statistical data on SP hardware and software and monitors system performance. It simplifies run-time performance monitoring on a large number of nodes.

The PTPE function of PSSP collects and provides performance data for SP hardware and software through enhancements to the Performance Toolbox for the AIX (PTX) product, which is the preferred performance monitor for AIX systems. It allows PTX to monitor unique SP subsystems, such as VSD, SP Switch, and LoadLeveler.

PTPE organizes the SP into a set of performance reporting groups with coordinating managers and distributes the burden of monitoring nodes throughout the SP system, thus, eliminating the need for a dedicated monitoring node. PTPE also provides average performance statistics for the SP system rather than monitoring every data point on all SP nodes. This can help reduce the computational effort required for run-time monitoring of SP performance.

Application programming models

PSSP supports a multi-threaded, standards-compliant Message Passing Interface (MPI) through an IBM Parallel Environment for AIX (PE) as well as maintaining its single-threaded MPI support. In addition, PSSP includes a Low-level Application Programming Interface (LAPI) with a flexible, active message style, communications programming model on the SP Switch.

10.2 General Parallel File System (GPFS)

This section describes General Parallel File System (GPFS).

10.2.1 Advantages

The following are the advantages of the GPFS:

- Provides a high-performance, scalable file system
- Recovers from most component failures
- Allows network access via NFS export
- Complies with UNIX file standards

- Allows flexible configuration
- Increases file I/O performance and scales with the system
- Optimizes SP Switch and disk utilization and supports multiple file block sizes
- Runs most AIX and UNIX applications unmodified
- Provides greater data availability and supports logging, replication, and component failover

10.2.2 Description

IBM General Parallel File System (GPFS) is a standards-based, parallel file system that delivers high performance, high availability, and high scalability while preserving the application interfaces used in standard file systems. GPFS allows access to files within an SP system from any GPFS node in the system and can be exploited by parallel jobs running on multiple SP nodes as well as by serial applications that are scheduled to nodes based on processor availability.

Background

Most UNIX file systems today run best in a single-server environment. Adding additional file servers does not necessarily improve the file access performance because the file systems were not designed to scale in a parallel processing environment. IBM introduced Parallel Input Output File System (PIOFS) to speed file access on the SP, but PIOFS is limited in usage due to its lack of recoverability. GPFS is the product that SP file users were waiting for. It was designed expressly to deliver scalable performance across multiple file system nodes, to comply with UNIX file standards, and to be recoverable from most failures. It delivers all of these functions.

Administration

GPFS provides functions that simplify multinode administration and can be performed from any node in the SP configuration. These functions are based on, and are in addition to, the AIX administrative commands that continue to operate. A single GPFS multinode command can perform a file system function across the entire SP system. In addition, most existing UNIX utilities will also run unchanged. All of these capabilities allow GPFS to be used as a replacement for existing UNIX file systems where parallel optimization is desired.

Standards compliance

GPFS supports the file system standards of X/OpenR 4.0, with minor exceptions, allowing most AIX and UNIX applications to use GPFS data without modification.

Higher performance/scalability

By delivering file performance across multiple nodes and disks, GPFS scales beyond single-server (node) performance limits. This level of performance is achieved through the use of IBM VSD, client-side data caching, large file block support, and the ability to perform read-ahead and write-behind file functions. As a result, GPFS can outperform PIOFS, Network File System (NFS), Distributed File System (DFS), and Journalled File System (JFS) in a parallel environment. Unlike NFS and JFS, GPFS file performance scales as additional file server nodes and disks are added to the SP system.

Availability/recoverability

GPFS can survive many system and I/O failures. Through its use of the RS/6000 Cluster Technology capabilities of PSSP in combination with IBM RVSD, GPFS is able to automatically recover from node, disk connection and disk adapter failures. GPFS will transparently failover lock servers and other GPFS central services. Through its use of IBM RVSD, GPFS continues to operate in the event of disk connection failures. GPFS allows data replication to further reduce the chances of losing data if storage media fail. Unlike PIOFS, GPFS is a logging file system that allows the recreation of consistent structures for quicker recovery after node failures. GPFS also provides the capability to mount multiple file systems, each of which can have its own recovery scope in the event of component failures.

Migration

Upgrading to a new release of GPFS can be tested on a GPFS configuration. This eases migration by allowing the testing of a new level of code without inhibiting the production GPFS application.

Why choose GPFS?

GPFS offers greater file performance than existing NFS, DFS, and PIOFS file systems on SP systems. It provides the scalability needed as SP systems grow with additional processors and disks while maintaining standard UNIX file interfaces. Finally, GPFS offers recoverability allowing it to replace most other file systems used on the SP today.

New in GPFS 1.3

GPFS Release 3 provides several performance, scalability, standards compliance, and usability enhancements. This includes improved performance in the areas of directory operations, strided I/O, administration

commands, and scalability. In addition, GPFS Release 3 offers exploitation of SP Security Services using Distributed Computing Environment (DCE) and export of GPFS file systems through Distributed File Service (DFS).

10.3 LoadLeveler

This section describes the LoadLeveler.

10.3.1 Advantages

The following are the advantages of the LoadLeveler:

- Distributed, full-function job scheduler
- Serial/parallel batch and interactive workload balancing
- Central point of control for workload administration
- Full scalability across processors and jobs
- API to enable alternate scheduling algorithms
- Single point of control provides job management and more effective workload scheduling
- Supports thousands of jobs across hundreds of SP nodes and IBM and non-IBM workstations
- Provides automatic recovery of central scheduler and can be configured with HACMP for node and network failover
- The primary job management solution for the RS/6000 family

10.3.2 Description

LoadLeveler is a distributed network-wide job management program for dynamically scheduling work on IBM and non-IBM processors. Sharing hardware resources has become common in the UNIX world. This is why LoadLeveler can help you get the most from your hardware investment. In the typical network, some processor nodes are overworked while others are idle. Valuable resources can be left unused especially during off-hours. LoadLeveler balances your workload, efficiently managing job flow in the network and distributing work across all LoadLeveler-defined hardware resources.

In a network of UNIX workstations or SP systems, LoadLeveler matches processing needs to available resources for improved performance and faster turnaround. LoadLeveler is an application that runs as a set of daemons on each IBM RS/6000 workstation and each SP node in the network. It provides

a facility for building, submitting, and processing batch jobs quickly and efficiently in a dynamic environment with a non-predictive job arrival pattern. For greater efficiency on an SP system, LoadLeveler with the Communication Subsystems Support (CSS) function of PSSP V3.1 can support up to four user space tasks per SP switch adapter.

Interfaces

LoadLeveler has a command line interface and a Motif-based graphical user interface (GUI) making it easy to submit and cancel jobs, monitor job status, and set and change job priorities. The system can be configured at the network and node levels, where workstations or SP nodes may be identified as job submitters, compute servers, or both. When a job is scheduled, its requirements are compared with all the resources known to LoadLeveler. Job requirements might be a combination of memory, disk space, architecture operating system, and application programs. LoadLeveler's Central Manager collects resource information and dispatches the job as soon as it locates suitable nodes.

LoadLeveler offers the option of using its own scheduler or an API to use alternative schedulers, such as EASY from the Cornell Theory Center. The product also provides a user- or system-initiated checkpoint/restart capability for certain types of FORTRAN, C, or C++ jobs linked to the LoadLeveler libraries.

What you can do with LoadLeveler

The LoadLeveler enables the following tasks for you:

- **Parallel processing.** For parallel jobs, LoadLeveler interfaces with parallel programming software, Parallel Environment for AIX, or PVM to obtain the multiple SP nodes required for the job's parallel tasks. In addition, APIs are available for linking to other parallel application environments. LoadLeveler V2.1 allows a task to run both the Message Passing Interface (MPI) and the Low-level Application Programming Interface (LAPI) communications protocols.
- **Individual control.** At the node level, users can specify to LoadLeveler when their processing nodes are available and how they are to be used. For example, some users might let their workstations accept any job during the night but only certain jobs during the day when they most need their resources. Other users might simply tell LoadLeveler to monitor their keyboard activity and make their workstations available whenever they have been idle for a sufficient time.
- **Central control.** From a system management perspective, LoadLeveler allows a system administrator to control all jobs running in a cluster

including the SP system. With hundreds of nodes configured, job and machine status are always available, thus providing administrators with the information needed to make adjustments to job classes and changes to LoadLeveler-controlled resources.

- Scalability. As nodes are added, LoadLeveler automatically scales upward so that the additional resources are transparent to the user.
- National language support (NLS). LoadLeveler is enabled for NLS. Error messages are externalized in both message catalogues and the LoadLeveler Diagnosis and Messages Guide.

New in LoadLeveler 2.2

Version 2.2 of LoadLeveler for AIX features a number of enhancements that increase flexibility while optimizing performance.

- More scheduling options:
Consumable resources allow users to schedule jobs based on the availability of specific resources.
- Improved process tracking:
LoadLeveler for AIX, Version 2.2 tracks processes spawned off by a job and cancels any processes left behind when a job is terminated, freeing up resources.
- Improved security:
By exploiting the PSSP Security Services within the Distributed Computing Environment (DCE), administrators can authenticate user identity and delegate credentials, so only authorized users can schedule jobs on specified resources.
- Faster job turnaround:
Version 2.2 incorporates a Process Tracking Function that locates processes leftover from terminated jobs. LoadLeveler can then cancel them and free up valuable resources for other tasks.
- More task assignment options:
Run an individual job on a specific node or assign tasks in blocks, let LoadLeveler determine the assignment, or use a combination of methods.
- Improved performance:
LoadLeveler daemons are now running multithreaded providing improved command response time.
- Improved ease of use:

The LoadLeveler Graphical User Interface (GUI) has been enhanced to provide a sequence of panels the user can step through to accomplish a task.

10.4 Parallel Engineering and Scientific Subroutine Library (PESSL)

This section describes the Parallel Engineering and Scientific Subroutine Library (PESSL).

10.4.1 Advantages

The following are the advantages of the PESSL:

- Provides mathematical algorithms optimized for high performance
- Supports RS/6000 SP systems and clusters of RS/6000 servers and/or workstations
- Callable from XL FORTRAN, C, and C++ applications
- Designed for high mathematical computational performance
- Supports many scientific and engineering applications used by multiple industries
- Tuned to the characteristics of RS/6000 hardware
- Can be used with existing programs via relinking rather than recompiling
- Supports easy development of parallel applications with Single Program Multiple Data model and/or Shared Memory Parallel Processing model for SMPs

10.4.2 Description

The Engineering and Scientific Subroutine Library (ESSL) family of products is a state-of-the-art collection of mathematical subroutines that provides a wide range of high-performance mathematical functions for many different scientific and engineering applications.

The ESSL family consists of:

- Engineering and Scientific Subroutine Library (ESSL) for AIX, which contains over 400 high-performance mathematical subroutines tuned to RS/6000 hardware. ESSL runs on RS/6000 workstations, servers, and SP systems.
- Parallel Engineering and Scientific Subroutine Library (PESSL) for AIX, which is specifically tuned to exploit the full power of the SP hardware with

scalability across the range of system configurations. In addition to SP systems, PESSL runs on clusters of RS/6000 servers and/or workstations.

ESSL and PESSL can be used to develop and enable many different types of scientific and engineering applications. New applications can be designed to take advantage of all the capabilities of ESSL family. Existing applications can be easily enabled by replacing comparable routines and in-line code with calls to ESSL subroutines.

Wide range of mathematical functions

ESSL provides a variety of complex mathematical functions, such as:

- Basic Linear Algebra Subroutines (BLAS)
- Linear Algebraic Equations
- Eigensystem Analysis
- Fourier Transforms

Examples of applications that use these types of mathematical subroutines are:

- Structural analysis
- Time series analysis
- Computational chemistry
- Computational techniques
- Fluid dynamics analysis
- Mathematical analysis
- Seismic analysis
- Dynamic systems simulation
- Reservoir modeling
- Nuclear engineering
- Quantitative analysis
- Electronic circuit design

The ESSL products are compatible with public domain subroutine libraries, such as Basic Linear Algebra Subprograms (BLAS), Scalable Linear Algebra Package (ScaLAPACK), and Parallel Basic Linear Algebra Subprograms (PBLAS), making it easy to migrate applications that utilize these libraries to use ESSL and/or Parallel ESSL.

What is new for ESSL Version 3 Release 2?

This section summarizes all the changes made to ESSL for AIX:

- The ESSL Libraries are tuned for the RS/6000 POWER3-II.
- The Dense Linear Algebraic Subroutines now include these new subroutines:
 - Symmetric Indefinite Matrix Factorization and Multiple Right-Hand Side Solve
 - Symmetric Indefinite Matrix Factorization
 - Symmetric Indefinite Matrix Multiple Right-Hand Side Solve
- The Linear Least Squares Subroutines now include this new subroutine:
 - General Matrix QR Factorization
- The ESSL POWER and Thread-Safe libraries have been replaced by a thread-safe library referred to as the ESSL Serial Library.
- The ESSL POWER2 and Thread-Safe POWER2 libraries are no longer provided; the ESSL Serial or the ESSL SMP Library should be used instead.

What is new for PESSL Version 2 Release 2?

This section summarizes all the changes made to PESSL for AIX:

- The Parallel ESSL Libraries are tuned for the RS/6000 POWER3-II Thin, Wide, and High nodes and the SP Switch2.
- The Dense Linear Algebraic Subroutines now include these new subroutines:
 - Inverse of a real or complex general matrix
 - Reciprocal of the condition number of a real or complex general matrix
 - General Matrix QR factorization
 - Least Squares solutions to linear systems of equations for real general matrices
- The Eigensystems Analysis Subroutines now include these new subroutines:
 - Selected Eigenvalues and optionally the eigenvectors of a real symmetric positive definite generalized eigenproblem
 - Reduce a complex Hermitian matrix to tridiagonal form
 - Reduce a real symmetric positive definite generalized eigenproblem to standard form

- The Utilities Subroutine now includes these new subroutines:
 - Compute the norm of a real or complex general matrix
- The PESSL POWER2 and Thread-Tolerant POWER2 libraries are no longer provided; the Parallel ESSL Serial and the Parallel ESSL SMP libraries should be used instead.
- Support is withdrawn for calling Parallel ESSL from HPF; as a result, the Parallel ESSL HPF libraries, HPF module, HPF IVP, and sample HPF programs are no longer provided.

For more detailed information on the Engineering and Scientific Subroutine Library (ESSL), and Parallel Engineering and Scientific Subroutine Library (PESSL), refer to the ESSL and PESSL manuals located at

http://www.rs6000.ibm.com/resource/aix_resource/sp_books/essl/index.html

10.5 Parallel Optimization Subroutine Library (OSLp)

This section describes the Parallel Optimization Subroutine Library (OSLp).

10.5.1 Advantages

The following are the advantages of the OSLp:

- Solves many types of mathematical programming problems:
 - Linear programming
 - Network programming
 - Mixed integer programming
 - Quadratic programming
- Contains all functions of AIX OSL/6000
- Used across multiple industries:
 - Transportation
 - Petroleum
 - Manufacturing
 - Finance
- Takes advantage of parallel processing to improve performance on mixed-integer and linear programming problems

10.5.2 Description

Parallel Optimization Subroutine Library (OSLp) is a collection of high performance mathematical subroutines used by application programmers to solve large optimization problems. It includes all of the functions of the AIX OSL/6000 product but solves linear and mixed-integer programming problems in parallel on the RS/6000 SP processor achieving a significant performance improvement.

Mathematical programming techniques can be applied to problems where the user wants to minimize or maximize an objective function subject to a set of constraints. A feasible solution solves the constraints; an optimal solution yields the largest or smallest value of the objective function among all feasible solutions.

OSLp is a set of over 60 subroutines callable from your application program in order to find the optimal solution to several types of problems using linear programming (LP), mixed-integer programming (MIP), and quadratic programming (QP) mathematical techniques. Some of the solutions use serial algorithms; that is, all computations are performed in sequence on an RS/6000 or single node of an SP system. Others use algorithms, which exploit the parallel processing capabilities of the SP system so that multiple nodes may concurrently perform computations on subtasks of the problem to be solved.

Parallel OSL models may be defined using OSL data structures, Mathematical Programming System (MPS) format for compatibility with predecessor IBM mathematical programming products, or Lotus 1-2-3 spreadsheet format. Subroutines are provided for loading models in any of these formats.

Operating environments

Parallel OSL is functionally equivalent to AIX OSL/6000, which should be used to solve mathematical programming problems when only a single RS/6000 processor is installed. However, taking advantage of multiple RS/6000s or SP nodes using OSL can solve selected optimization problems more rapidly.

Parallel OSL differs from OSL/6000 by virtue of the replacement of two major solver subroutines, EKKMSLV for interior point LP and EKKBSLV for MIP problems. Only minor changes to serial OSL applications are required to generate parallel application programs.

No explicit parallel coding is required. The parallel solvers have the same names and calling sequences as their serial counterparts. Access to OSL

parallel processing is provided by new parameter values and a few new control variables. Parallel OSL's mathematical subroutines are callable from user application programs written in XL Fortran or C. High-level subroutines can solve a problem with the user having minimal knowledge of mathematical programming. Low-level subroutines give the user the flexibility to structure algorithms without having to write new routines independently.

A supported parallel execution environment must be present to execute Parallel OSL on concurrent parallel processors. These separate products provide the environment for the execution in parallel of subproblems of the mathematical optimization including allocation of multiple processors and invocation and monitoring of subproblem execution.

Linear programming (LP)

In an LP, both the objective function and the constraints are linear. The several algorithms available in Parallel OSL are:

- Simplex method - OSL uses either a primal or a dual simplex serial algorithm.
- Interior Point Barrier method - OSL uses the primal barrier, primal-dual barrier, or primal-dual with predictor-corrector fine-grained parallelized algorithms.
- Network Solver method - OSL solves this special case using a serial algorithm.

Mixed-integer programming (MIP)

MIP problems are LPs in which some variables are constrained to be integers. Parallel OSL includes a versatile, course-grain parallelized branch and bound solver handling MIP problems with either linear or quadratic objective functions. The simplex solver (primal or dual, selectable) is used on LP sub-problems; the QP solver is used on QP sub-problems. You may use as many SP nodes as are available (and licensed for OSL). These nodes may be used to process different branches of the search tree and to improve the performance of the algorithm.

Quadratic programming (QP)

QP problems have a convex quadratic objective function and linear constraints. Parallel OSL includes a fast two-stage serial algorithm for solving these problems. The first sub-algorithm solves an approximating LP problem and a related very simple QP problem at each iteration. When successive approximations are close enough together, the second sub-algorithm is used. This extension of the simplex method permits a quadratic objective function and converges very rapidly when given a good starting value.

Other solution capabilities

Subroutines are provided for sensitivity analysis allowing evaluation of the effect of objective function coefficient changes on the optimal solution basis. Also provided is a parametric solver that shows how the objective value and optimal solution vary as row bounds, column bounds, and objective function coefficients change over a range.

Performance

Benchmarks have shown that Parallel OSL achieves significant speedups on LP and MIP problems with its parallelized solution algorithms. LPs solved with the interior point algorithm are typically able to achieve a sub-linear speedup of 40 to 50 percent processor utilization. That is, running such a problem on an SP with eight processors is three times as fast as running OSL/6000 on a single node.

MIP problems perform even better on a parallel processor because the MIP branch-and-bound solution algorithm requires little interprocess communication so that multiple processors can operate concurrently and independently with little need for synchronization. Nearly linear speedup is regularly achieved with occasional superlinear speedup. Processor utilization is typically in the 90 to 100 percent range resulting in a speedup of seven to eight times faster on an eight processor SP.

Features and functions

The following are the features and functions:

- **Compatible with OSL** - Functionally-equivalent and source-compatible to the AIX OSL/6000 product.
- **Solution economies** - Provides sophisticated methods that prepare a problem prior to submission for solution by producing a basis quickly, thereby, eliminating redundant constraints to reduce problem size and scaling the coefficient matrix.
- **Control variables** - Variables accessible by the application programmer to fine tune solution algorithm.
- **Debugging capabilities** - Routines available to provide statistics about element values and storage used during execution.
- **MPS format compatibility** - Accepts industry standard MPS input format for compatibility with previous mathematical programming solvers, such as MPSX/370.
- **Spreadsheet format** - Accepts as input models in Lotus 1-2-3 spreadsheet format for ease of data maintenance.

- **Utilities** - Provide flexibility and productivity by allowing alteration of matrices, variation of print formats, tailoring of user exits, and model debugging.
- **User exits** - Exits allow application programs to gain information/control at every iteration of solve subroutine or when message is issued. Used to record statistics, report progress, intercept messages, or stop execution.

Mixed-integer exits are initially used for diagnosing what solution choices are being made. Exits may then be used to alter the integer solution algorithm.

10.6 Parallel Environment (PE)

This section describes the Parallel Environment (PE).

The following are the advantages of the PE:

- Provides a development and execution environment for parallel applications
- Exploits threads and thread-safe MPI message passing on all nodes including symmetric multiprocessors (SMPs)
- Supports Low-level Application Programming Interface (LAPI) programs
- Easier parallel application development
- Enhanced XProfiler graphical performance tool
- Enables easy application portability to networked RS/6000 or RS/6000 SP systems

10.6.1 Description

Parallel Environment (PE) for AIX on the RS/6000 platform is a complete solution for enterprises that need to develop, debug, analyze, tune, and execute parallel programs on the AIX platform.

This application development solution consists of:

- Parallel Message Passing APIs for full implementation of the MPI 1.2 standard plus full implementation of the MPI-I/O and MPI-1 sided communication chapters of the MPI-2 standards. Also, continued support for the IBM Message Passing Library for communications between executing tasks in a Fortran, C, or C++ parallel program.
- A Parallel Operating Environment (POE) for managing the development and execution of parallel applications.

- Parallel debuggers offering both command-line and Motif-based interfaces. These debuggers extend traditional AIX capabilities and provide features for parallel application task debugging.
- Xprofiler graphical performance tool
- Dynamic Probe Class Library (DPCL) parallel tools development API

PE V2.3 functions

PE V2.3 provides exploitation of threads and thread-safe MPI message passing on all nodes, including SMPs, and also supports Low-level Application Programming Interface (LAPI) programs. Included are:

- Tool extension for threaded applications
- Support for parallel applications on AIX 4.2.1
- Easy application portability to a networked cluster of RS/6000 or RS/6000 SP systems

The POE has been enhanced to support threaded applications and to handle asynchronous signals from the user parallel program. The visualization and performance monitoring tool has been enhanced to support thread-safe VT trace output files in addition to limited visualizations of thread events. A graphical tool, XProfiler, is available for analyzing application performance. Zoom, filter, and search functions support graphical manipulation to aid analysis and tuning of parallel programs.

PE V2.4

PE V2.4 increases the number of tasks supported in a single POE job from 512 to 2048 (1024 with user space MPI or LAPI libraries). In addition, up to four user space tasks can now run on each SP node from the same or different jobs, thus allowing parallel applications to exploit SMP nodes. POE now also provides the ability to checkpoint the state of a parallel batch program for restart in case of application failure.

For increased MPI application debugging, a new PEDB capability provides the following:

- A summary of the number of active messages from each task in an application
- Message queue information for a specific task
- Detailed confirmation for a specific message
- The ability to debug parallel applications that are currently executing
- The ability to view thread events in the visualization and performance monitoring tool

Why choose PE?

PE is the IBM strategic high-function development and execution environment for parallel applications using either the RS/6000 SP system or one or more RS/6000 processors.

PE has been enhanced to exploit SMP nodes as well as threaded applications and the MPI message passing API.

New in PE 3.1

The IBM Parallel Environment for AIX Version 3, Release 1 is a high-function development and execution environment for parallel applications using either the RS/6000 SP system or one or more RS/6000 processors. This newest version supports AIX V4.3.3 operating system and includes the following function enhancements and changes:

- New application program interfaces (APIs)
- Dynamic probe class library (DPCL) parallel tools development API
This version of Parallel Environment for AIX V3.1(PE 3.1) includes a new set of interfaces called the dynamic probe class library (DPCL). With DPCL, tool builders can define instrumentation that can be inserted and removed from an application as it is running. Because the task of generating instrumentation is simplified, designers can develop new tools quickly and easily using DPCL.
- Parallel task identification API
This version of PE V3.1 introduces full support for the new Parallel Operating Environment (POE) API that allows an application to retrieve the process IDs of all POE master processes running on the same node. This information can be used for accounting or to get more detailed information about the tasks spawned by these POE processes.
- New support for FORTRAN 95
PE 3.1 now additionally supports FORTRAN 95.
- Support for Distributed Computing Environment (DCE) security
This version of PE introduces full support for the Distributed Computing Environment (DCE) through the SP Security Services of PSSP. The use of DCE security is optional.
- MPI enhancements
 - Full support for MPI I/O
PE 3.1 now provides full support for MPI I/O interfaces.
 - Support for MPI one-sided communication

With this release, PE now provides full support for MPI one-sided communication that allows one process to specify all communication parameters for the sending operation as well as the receiving operation.

- Support for MPI shared memory message passing

PE 3.1 introduces support for shared memory MPI message passing on symmetric multiprocessor (SMP) nodes, for the Internet Protocol (IP) library and for the User Space (US) library. This support includes a new environment variable that lets you select the shared memory protocol. MPI programs may benefit from using shared memory to send messages between two or more tasks that are running on the same node reducing adapter communication traffic. Your applications do not need to be changed in any way to take advantage of this support.

- Support for 4096 tasks

With this version, PE 3.1 increases the maximum number of user space (US) tasks from 1024 to 4096 if the hardware supports this size job. For SP Switch Systems, the hardware supports a maximum of 4 tasks per node, with up to 512 nodes, for a total of 2048 User Space tasks. For SP Switch2 Systems, the hardware supports a maximum of 16 tasks per node, with up to 512 nodes, for a total of 8192, but PE only supports up to 4096 task jobs. The PEDB debugger supports a maximum of 32 tasks. The IP limits are still 2048 tasks.

- Removal of Parallel Environment Version 1 support

Beginning with this release, PE V.1 is no longer supported.

- Removal of VT support

Beginning with this release, PE no longer includes the visualization tool (VT) function.

10.7 Performance Toolbox Parallel Extensions (PTPE)

This section describes the Performance Toolbox Parallel Extensions (PTPE) feature.

10.7.1 Advantages

The following are the advantages of the PTPE:

- Extends the Performance Toolbox for the AIX product to the SP environment.
- Lightens administrative loads.

- Provides run-time monitoring.
- Records performance data and saves it for later analysis.
- Familiar run-time displays show how RS/6000 SP nodes are running.
- Monitor as many aspects of system performance as you want on as many SP nodes as you want. You can group nodes for meaningful summary statistics and display the results now or archive them for later analysis.
- Distribute data management across multiple nodes, eliminating bottlenecks and preventing your performance analysis effort from affecting system performance.
- Collect and display statistical data for SP hardware (nodes and switches) and IBM software (LPPs), such as LoadLeveler and PSSP.

10.7.2 Description

PTPE is a feature of PSSP. PSSP is a collection of administrative and operational software applications that enables system administrators and operators to more easily manage RS/6000 SP systems and their environments.

PTPE simplifies performance analysis and reduces administrative overhead by organizing your SP nodes into reporting groups. Each node sends performance data to a manager node, which performs the administrative tasks for the group, calculating averages and retrieving data. This shared responsibility among manager nodes prevents the monitoring effort from hampering system performance.

You can have PTPE group your SP nodes automatically, or you can define customized reporting groups that yield more meaningful performance summaries. It might be useful to group nodes logically according to the kind of tasks they perform. For example, interactive nodes might form one group, while batch processing nodes might belong to another.

Of course, you can always display performance data for a single node, but you will soon come to appreciate the convenience of analyzing multiple nodes with one statistic.

Archive statistics

Analysis of historical data is sometimes the only way performance trends become apparent; so, you might want to track performance over a period of time without displaying it now. Although run-time monitors display only a snapshot of system performance, the time span and range of performance data that PTPE can archive are limited only by your system's storage

capacity. You can tell PTPE to collect and store for future use, in machine-readable format, any statistics generated by your operating system or IBM LPPs.

Archival is completely independent of the run-time monitor, so the data you save for later need not be the same data you are collecting and displaying for current performance analysis. Archived performance data is available at any time for print or export to database or spreadsheet applications.

Precise control

You can start, stop, and control the operation of PTPE through the use of simple AIX commands or by using the performance monitoring facility included in the SP Perspectives.

Much of the PTPE setup and configuration can be performed automatically. You can allow PTPE to group your SP nodes by frame. After reviewing the statistics PTPE collected, you might want to group your nodes differently or disable some of the performance data collected by default.

Application programming library

PTPE includes a set of subroutines for use in your own application programs. These subroutines enable you to tailor the collection, archival, and extraction of performance statistics in a number of ways, therefore, making PTPE a versatile, yet easy-to-use, tool for SP system administration.

Start monitoring now

Make the most of your SP's computing power. Use PTPE to monitor node performance so that you can balance your SP workload and tune your configuration for maximum throughput. Whether you are interested in run-time analysis or compiling a performance history, PTPE can give you easy access to all the performance data available on your SP system.

Appendix A. Special notices

This publication is intended to help IBM sales professionals, Business Partners, ISVs, and customers wishing to obtain a reference for IBM RS/6000 SP system hardware and software offerings. The information in this publication is not intended as the specification of any programming interfaces that are provided by AIX software or by RS/6000 SP system hardware and software. See the Publications section of the IBM Programming Announcement for RS/6000 SP system and AIX LPPs for more information about what publications are considered to be product documentation.

References in this publication to IBM products, programs or services do not imply that IBM intends to make these available in all countries in which IBM operates. Any reference to an IBM product, program, or service is not intended to state or imply that only IBM's product, program, or service may be used. Any functionally equivalent program that does not infringe any of IBM's intellectual property rights may be used instead of the IBM product, program or service.

Information in this book was developed in conjunction with use of the equipment specified, and is limited in application to those specific hardware and software products and levels.

IBM may have patents or pending patent applications covering subject matter 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 the IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785.

Licensees of this program who wish to have information about it for the purpose of enabling: (i) the exchange of information between independently created programs and other programs (including this one) and (ii) the mutual use of the information which has been exchanged, should contact IBM Corporation, Dept. 600A, Mail Drop 1329, Somers, NY 10589 USA.


Such information may be available, subject to appropriate terms and conditions, including in some cases, payment of a fee.

The information contained in this document has not been submitted to any formal IBM test and is distributed AS IS. The use of this information or the implementation of any of these techniques is a customer responsibility and depends on the customer's ability to evaluate and integrate them into the customer's operational environment. While each item may have been reviewed by IBM for accuracy in a specific situation, there is no guarantee

that the same or similar results will be obtained elsewhere. Customers attempting to adapt these techniques to their own environments do so at their own risk.

Any pointers in this publication to external Web sites are provided for convenience only and do not in any manner serve as an endorsement of these Web sites.

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

ADSTAR	AIX
AIX/ESA	
CT	DB2
DirectTalk	ES/9000
ESCON	IBM
IMS	LoadLeveler
Magstar	Micro Channel
MVS/ESA	OS/390
Portmaster	PowerPC 604
Redbooks	Redbooks Logo 
RS/6000	S/390
SP	SP1
SP2	System/390
TURBOWAYS	VM/ESA

The following terms are trademarks of other companies:

Tivoli, Manage. Anything. Anywhere., The Power To Manage., Anything. Anywhere., TME, NetView, Cross-Site, Tivoli Ready, Tivoli Certified, Planet Tivoli, and Tivoli Enterprise are trademarks or registered trademarks of Tivoli Systems Inc., an IBM company, in the United States, other countries, or both. In Denmark, Tivoli is a trademark licensed from Kjøbenhavns Sommer - Tivoli A/S.

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

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

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

PC Direct is a trademark of Ziff Communications Company in the United

States and/or other countries and is used by IBM Corporation under license.

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

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

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

Lotus Notes is a registered trademark of Lotus Development Corporation.

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

Appendix B. Related publications

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

B.1 IBM Redbooks

For information on ordering these publications see “How to get IBM Redbooks” on page 211.

- *GPFS: A Parallel File System*, SG24-5165
- *HACMP Enhanced Scalability Handbook*, SG24-5328
- *HACMP Enhanced Scalability User-Defined Events*, SG24-5327
- *IBM 9077 SP Switch Router: Get Connected to the SP Switch*, SG24-5157
- *IBM Certification Study Guide: RS/6000 SP*, SG24-5348
- *RS/6000 SP Software Maintenance*, SG24-5160
- *RS/6000 SP System Performance Tuning*, SG24-5340
- *SP Perspectives: A New View of Your SP System*, SG24-5180
- *The RS/6000 SP Inside Out*, SG24-5374
- *Understanding and Using the SP Switch*, SG24-5161

B.2 IBM Redbooks collections

Redbooks are also available on the following CD-ROMs. Click the CD-ROMs button at ibm.com/redbooks for information about all the CD-ROMs offered, updates and formats.

CD-ROM Title	Collection Kit Number
IBM System/390 Redbooks Collection	SK2T-2177
IBM Networking Redbooks Collection	SK2T-6022
IBM Transaction Processing and Data Management Redbooks Collection	SK2T-8038
IBM Lotus Redbooks Collection	SK2T-8039
Tivoli Redbooks Collection	SK2T-8044
IBM AS/400 Redbooks Collection	SK2T-2849
IBM Netfinity Hardware and Software Redbooks Collection	SK2T-8046
IBM RS/6000 Redbooks Collection	SK2T-8043
IBM Application Development Redbooks Collection	SK2T-8037
IBM Enterprise Storage and Systems Management Solutions	SK3T-3694

B.3 Other resources

These publications are also relevant as further information sources:

- *RS/6000 SP: Maintenance Information, Volume 1, Installation and Relocation*, GA22-7375
- *RS/6000 SP: Maintenance Information, Volume 2, Maintenance Analysis Procedures*, GA22-7376
- *RS/6000 SP: Maintenance Information, Volume 3, Locations and Service Procedures*, GA22-7377
- *RS/6000 SP: Maintenance Information, Volume 4, Parts Catalog*, GA22-7378
- *RS/6000 SP: Planning, Volume 1, Hardware and Physical Environment*, GA22-7280
- *RS/6000 SP: Planning, Volume 2, Control Workstation and Software Environment*, GA22-7281

B.4 Referenced Web sites

This Web site is also relevant as a further information source:

- <http://www.rs6000.ibm.com/hardware/largescale/index.html>

How to get IBM Redbooks

This section explains how both customers and IBM employees can find out about IBM Redbooks, redpieces, and CD-ROMs. A form for ordering books and CD-ROMs by fax or e-mail is also provided.

- **Redbooks Web Site** ibm.com/redbooks

Search for, view, download, or order hardcopy/CD-ROM Redbooks from the Redbooks Web site. Also read redpieces and download additional materials (code samples or diskette/CD-ROM images) from this Redbooks site.

Redpieces are Redbooks in progress; not all Redbooks become redpieces and sometimes just a few chapters will be published this way. The intent is to get the information out much quicker than the formal publishing process allows.

- **E-mail Orders**

Send orders by e-mail including information from the IBM Redbooks fax order form to:

	e-mail address
In United States or Canada	pubscan@us.ibm.com
Outside North America	Contact information is in the "How to Order" section at this site: http://www.elink.ibm.com/pbl/pbl

- **Telephone Orders**

United States (toll free)	1-800-879-2755
Canada (toll free)	1-800-IBM-4YOU
Outside North America	Country coordinator phone number is in the "How to Order" section at this site: http://www.elink.ibm.com/pbl/pbl

- **Fax Orders**

United States (toll free)	1-800-445-9269
Canada	1-403-267-4455
Outside North America	Fax phone number is in the "How to Order" section at this site: http://www.elink.ibm.com/pbl/pbl

This information was current at the time of publication, but is continually subject to change. The latest information may be found at the Redbooks Web site.

IBM Intranet for Employees

IBM employees may register for information on workshops, residencies, and Redbooks by accessing the IBM Intranet Web site at <http://w3.itso.ibm.com/> and clicking the ITSO Mailing List button. Look in the Materials repository for workshops, presentations, papers, and Web pages developed and written by the ITSO technical professionals; click the Additional Materials button. Employees may access MyNews at <http://w3.ibm.com/> for redbook, residency, and workshop announcements.

List of Abbreviations

AC	Alternating Current	CSU	Channel Service Unit
ADSM	ADSTAR Distributed Storage Manager	CWS	Control Workstation
ADSTAR	Advanced Storage and Retrieval	DAS	Dual Attach Station
AIX	Advanced Interactive Executive	DASD	Direct Access Storage Device (Disk)
ANSI	American National Standards Institute	DC	Direct Current
APAR	Authorized Program Analysis Report	DE	Dual-Ended
API	Application Programming Interface	DFS	Distributed File System
ASCI	Accelerated Strategic Computing Initiative	DIMM	Dual In-Line Memory Module
ASCII	American National Standards Code for Information Interchange	DMA	Direct Memory Access
ATM	Asynchronous Transfer Mode	DRAM	Dynamic Random Access Memory
BLAS	Basic Linear Algebra Subprograms	DSU	Data Service Unit
BOS	Base Operating System	ECC	Error Checking and Correction
CD-ROM	Compact Disk-Read Only Memory	EIA	Electronics Industry Association
CEC	Central Electronics Complex	EMIF	ESCON Multiple Image Facility
CLIO/S	Client Input/Output Sockets	ESCON	Enterprise Systems Connection (Architecture, IBM System/390)
CMOS	Complimentary Metal Oxide Semiconductor	ESSL	Engineering and Scientific Subroutine Library
CPU	Central Processing Unit	F/C	Feature Code
CSMA/CD	Carrier Sense Multiple Access/Collision Detection	FDDI	Fiber Distributed Data Interface
CSS	Communication Subsystems Support	FTP	File Transfer Protocol
		F/W	Fast and Wide
		GPFS	General Parallel File System
		GUI	Graphical User Interface

HACMP	High Availability Cluster Multi-Processing	LAN	Local Area Network
HACWS	High Availability Control Workstation	LANE	Local Area Network Emulation
HIPPI	High Performance Parallel Interface	LAPI	Low-Level Application Programming Interface
HIPS	High Performance Switch	LED	Light Emitting Diode
HIPS LC-8	Low-Cost Eight-Port High Performance Switch	LP	Linear Programming
HP	Hewlett-Packard	LPP	Licensed Program Product
HPF	High Performance FORTRAN	MAU	Multiple Access Unit
HPSSDL	High Performance Supercomputer Systems Development Laboratory	Mbps	Megabits Per Second
Hz	Hertz	MBps	Megabytes Per Second
IBM	International Business Machines Corporation	MCA	Micro Channel Architecture
IEEE	Institute of Electrical and Electronics Engineers	MES	Miscellaneous Equipment Specification
I/O	Input/Output	MIP	Mixed-Integer Programming
IP	Internetwork Protocol (OSI)	MMF	Multi-Mode Fiber
IPL	Initial Program Load	MP	Multiprocessor
ISA	Industry Standard Architecture	MP	Multi-Purpose
ISB	Intermediate Switch Board	MPI	Message Passing Interface
ISV	Independent Software Vendor	MPP	Massively Parallel Processing
ITSO	International Technical Support Organization	MPS	Mathematical Programming System
JFS	Journalled File System	MVS	Multiple Virtual Storage (IBM System 370 and 390)
L1	Level 1	MX	Mezzanine Bus
L2	Level 2	NFS	Network File System
		NIM	Network Installation Manager
		NTP	Network Time Protocol
		NVRAM	Non-Volatile Random Access Memory

OS/390	Operating System/390	RDBMS	Relational Database Management System
OSL	Optimization Subroutine Library	RPA	RS/6000 Platform Architecture
OSLp	Parallel Optimization Subroutine Library	RVSD	Recoverable Virtual Shared Disk
P2SC	Power2 Super Chip	SAS	Single Attach Station
PBLAS	Parallel Basic Linear Algebra Subprograms	ScaLAPACK	Scalable Linear Algebra Package
PCI	Peripheral Component Interconnect	SCSI	Small Computer System Interface
PE	Parallel Environment	SDR	System Data Repository
PEDB	Parallel Environment Debugging	SDRAM	Synchronous Dynamic Random Access Memory
PIOFS	Parallel Input Output File System	SDLC	Synchronous Data Link Control
POE	Parallel Operating Environment	SE	Single-Ended
POWER	Performance Optimization with Enhanced Risc (Architecture)	SEPBU	Scalable Electrical Power Base Unit
PSSP	Parallel System Support Programs	SMP	Symmetric Multiprocessor
PTF	Program Temporary Fix	SMT	Station Management
PTPE	Performance Toolbox Parallel Extensions	SNMP	Simple Network Management Protocol
PTX	Performance Toolbox	SP	Scalable POWERParallel
PVC	Permanent Virtual Circuit	SP	Service Processor
QP	Quadratic Programming	SPS	SP Switch
RAM	Random Access Memory	SPS-8	Eight-Port SP Switch
RAN	Remote Asynchronous Node	SSA	Serial Storage Architecture
RAS	Reliability, Availability, and Serviceability	STP	Shielded Twisted Pair
RAID	Redundant Array of Independent Disks	SUP	Software Update Protocol
		SVC	Switch Virtual Circuit
		Tcl	Tool Command Language

TCP/IP	Transmission Control Protocol/Internet Protocol
TPC	Transaction Processing Council
UDB EEE	Universal Database and Enterprise Extended Edition
UP	Uniprocessor
UTP	Unshielded Twisted Pair
VM	Virtual Machine (IBM System 370 and 390)
VSD	Virtual Shared Disk
VT	Visualization Tool
WAN	Wide Area Network

Index

Numerics

- 1.2 MB/sec Rack Mountable Remote Asynchronous Node, 16-port, EIA-232 147
- 1.2 MB/sec Remote Asynchronous Node, 16-port, EIA-232 (US) 147
- 1.2 MB/sec Remote Asynchronous Node, 16-port, EIA-232 (world trade) 147
- 10/100 Ethernet Twisted Pair MC Adapter 174
- 128-port Asynchronous Controller Cable 23 cm (9 in.) 147
- 128-port Asynchronous Controller Node Cable, 4.5 m 147
- 2.4 MB/sec Enhanced Remote Asynchronous Node, 16-port, EIA-232 147
- 2.4 MB/sec Enhanced Remote Asynchronous Node, 16-port, RS-422 147
- 2-port Multiprotocol X.25 Adapter 149
- 332 MHz SMP node 57
 - architecture 58
 - I/O bridge 63
 - I/O bridge bus 62
 - L2 cache 60
 - memory-I/O controller 61
 - microprocessor 60
 - system bus 59
- 332 MHz SMP thin node
 - bus 64
 - configuration rule 64
 - disk option 67
 - disk requirement 67
 - memory option 66
 - memory requirement 66
 - option 65
 - processor option 66
 - processor requirement 66
 - requirement 65
 - switch adapter option 67
 - switch adapter requirement 67
 - switch restriction 67
- 332 MHz SMP thin nodes 64
- 332 MHz SMP wide node
 - adapter placement restriction 68
 - bus 68
 - disk option 71
 - disk requirement 71
 - memory option 70
 - memory requirement 70
 - option 70
 - processor option 70
 - processor requirement 70
 - requirement 69
 - switch adapter option 72
 - switch adapter requirement 72
 - switch restriction 72
- 375 MHz POWER3 SMP Thin Node 44
- 375 MHz POWER3 SMP Wide Node 39
- 4-Port Multiprotocol Communications Controller 170
- 64 MB DRAM SIMM 179
- 7017-S70 Advanced Enterprise Server 73
- 7017-S70 Enterprise Server 73
- 7017-S7A Enterprise Server 73
- 7017-S80 Enterprise Server 73
- 8-Port Async Adapter - EIA-232 172
- 8-Port Async Adapter - EIA-422A 172
- 9333 High Performance Subsystem Adapter 174

A

- accounting 182
- AIX OSL/6000 195
- AIX/ESA 4
- ARTIC960Hx 4-port E1 RJ45 cable 168
- ARTIC960Hx 4-port EIA-232 cable 149
- ARTIC960Hx 4-port EIA-530 cable 149
- ARTIC960Hx 4-port RS-449 cable 149
- ARTIC960Hx 4-port Selectable Adapter 148
- ARTIC960Hx 4-port T1 RJ45 cable 168
- ARTIC960Hx 4-port V.35 (DTE) cable 149
- ARTIC960Hx 4-port X.21 cable 149
- ARTIC960RxD Quad Digital Trunk Adapter 167
- ARTIC960RxD Quad DTA, E1, 120 ohm balanced, 3 m 4-port cable 168
- ARTIC960RxD Quad DTA, E1, 120 ohm balanced, 7.5 m extension cable 168
- ARTIC960RxD Quad DTA, E1, 75 ohm unbalanced-grounded, 1.8 m 4-port cable 168
- ARTIC960RxD Quad DTA, E1, 75 ohm unbalanced-ungrounded, 1.8 m 4-port cable 168
- ARTIC960RxD Quad DTA, H.100, 4-drop cable 168
- ARTIC960RxD Quad DTA, T1, 100 ohm, 15 m extension cable 168
- ARTIC960RxD Quad DTA, T1, 100 ohm, 3 m 4-port

cable 168
Ascend 97
Ascend GRF 1600 97
Ascend GRF 400 97
Asynchronous Terminal/Printer Cable, EIA-232 147
ATM 155 MMF 159
ATM 155 TURBOWAYS UTP Adapter 151
ATM OC12, one port MM fiber 179
ATM OC12, one port SM fiber 178
ATM OC3, two port MM fiber 177
ATM OC3, two port SM fiber 177
Auto Token-Ring LANstreamer MC 32 Adapter 173

B

Basic Linear Algebra Subprograms 193
Blank faceplate 178
BLAS 193
Block Multiplexer Channel Adapter - BMCA 172
business intelligence 8

C

cable 13
Central Manager 190
Clustered Enterprise Server 91
Communication Subsystems Support 185, 190
control workstation 123
 interface adapter requirement 128
 software requirement 130
 supported IBM RS/6000 workstation 125
 MCA control workstation 126
 PCI control workstation 125
 system requirement 126
Cornell Theory Center 190
CSS 190

D

dependent node 13
dependent node adapter 13
DFS 188
Distributed File System 188
drawer 15

E

EASY 190
e-business 8
EIA 232/RS-422 8-Port Asynchronous Adapter 144

EIA 232D/V.24 cable 151
EKKBSLV 196
EKKMSLV 196
Engineering and Scientific Subroutine Library 192
Enhanced SCSI-2 Differential Fast/Wide Adapter/A 169
Enhanced SSA 4-Port Adapter 174
enterprise resource planning 8
ERP 8
ESCON Control Unit Adapter 172
ESSL 192
Ethernet 10 MB AUI 158
Ethernet 10 MB BNC 157
Ethernet 10/100 MB 153
Ethernet 10/100Base-T, eight port 178
Ethernet 10/100Base-T, four port 178
Ethernet 10BaseT Transceiver 174
Ethernet High Performance LAN Adapter 173
Ethernet LAN Adapter (AUI/10BaseT) 173
Ethernet LAN Adapter 10Base2 (BNC) 174
Event Management 184
extension node 13

F

FDDI Attachment dual-ring 171
FDDI Attachment single-ring 171
FDDI SK-NET LP DAS 139
FDDI SK-NET LP SAS 138
FDDI SK-NET UP DAS 141
FDDI, four port MM fiber 178
feature code
 1101 ATM OC3, two port SM fiber 177
 1102 ATM OC3, two port MM fiber 177
 1103 SONET/IP OC3, one port MM fiber 177
 1104 SONET/IP OC3, one port SM fiber 178
 1105 ATM OC12, one port SM fiber 178
 1106 FDDI, four port MM fiber 178
 1107 Ethernet 10/100Base-T, eight port 178
 1108 HIPPI, one port 178
 1109 HSSI, two port 178
 1112 Ethernet 10/100Base-T, four port 178
 1113 Blank faceplate 178
 1114 64 MB DRAM SIMM 179
 1115 ATM OC12, one port MM fiber 179
 1213 redundant power 17
 1241 Independent SCSI Hookup 71
 1500 short expansion frame 19
 1550 tall expansion frame 24

18.2 GB Ultra SCSI 10K RPM disk pair 38
 2050 332 MHz SMP thin node 64
 2051 332 MHz SMP wide node 67
 2054 POWER3 SMP High Node 52
 2056 375 MHz POWER3 SMP Thin Node 44
 2057 375 MHz POWER3 SMP Wide Node 39
 2058 375 MHz POWER3 SMP High Node 35
 2402 IBM Network Terminal Accelerator - 256
 Session 168
 2403 IBM Network Terminal Accelerator - 2048
 Session 169
 2410 SCSI-2 High Performance External I/O
 Controller 169
 2412 Enhanced SCSI-2 Differential Fast/Wide
 Adapter/A 169
 2415 SCSI-2 Fast/Wide Adapter/A 170
 2416 SCSI-2 Differential Fast/Wide Adapter/A
 170
 2420 SCSI-2 Differential External I/O Controller
 170
 2700 4-Port Multiprotocol Communications Con-
 troller 170
 2709 ARTIC960Hx 4-port T1 RJ45 cable 168
 2710 ARTIC960Hx 4-port E1 RJ45 cable 168
 2723 FDDI Attachment dual-ring 171
 2724 FDDI Attachment single-ring 171
 2735 HIPPI 171
 2741 FDDI SK-NET LP SAS 138
 2742 FDDI SK-NET LP DAS 139
 2743 FDDI SK-NET UP DAS 141
 2751 S/390 ESCON Channel Adapter 142
 2754 S/390 ESCON Channel Emulator Adapter
 171
 2755 Block Multiplexer Channel Adapter -
 BMCA 172
 2756 ESCON Control Unit Adapter 172
 2861 ARTIC960Hx 4-port EIA-232 cable 149
 2862 ARTIC960Hx 4-port RS-449 cable 149
 2863 ARTIC960Hx 4-port X.21 cable 149
 2864 ARTIC960Hx 4-port V.35 (DTE) cable
 149
 2865 ARTIC960Hx 4-port EIA-530 cable 149
 2871 ARTIC960RxD Quad DTA, T1, 100 ohm, 3
 m 4-port cable 168
 2872 ARTIC960RxD Quad DTA, T1, 100 ohm,
 15 m extension cable 168
 2873 ARTIC960RxD Quad DTA, E1, 120 ohm
 balanced, 3 m 4-port cable 168
 2874 ARTIC960RxD Quad DTA, E1, 120 ohm
 balanced, 7.5 m extension cable 168
 2875 ARTIC960RxD Quad DTA, E1, 75 ohm un-
 balanced-grounded, 1.8 m 4-port cable 168
 2876 ARTIC960RxD Quad DTA, E1, 75 ohm un-
 balanced-ungrounded, 1.8 m 4-port cable 168
 2877 ARTIC960RxD Quad DTA, H.100, 4-drop
 cable 168
 2900 4.5 GB Ultra SCSI disk drive 67, 71
 2904 4.5 GB Ultra SCSI disk drive pair 67, 71
 2908 9.1 GB Ultra SCSI disk drive 67, 71
 2909 9.1 GB Ultra SCSI disk drive pair 67, 71
 2918 18.2 GB Ultra SCSI disk drive pair 67, 71
 2920 Token Ring Auto LANstreamer 143
 2930 8-Port Async Adapter - EIA-232 172
 2934 Asynchronous Terminal/Printer Cable,
 EIA-232 147
 2940 8-Port Async Adapter - EIA-422A 172
 2943 EIA 232/RS-422 8-Port Asynchronous
 Adapter 144
 2944 WAN RS232 128-port 146
 2947 ARTIC960Hx 4-port Selectable Adapter
 148
 2951 EIA 232D/V.24 cable 151
 2952 V.35 cable 151
 2953 V.36/EIA 449 cable 151
 2954 X.21 cable 151
 2960 X.25 Interface Co-Processor/2 172
 2962 2-port Multiprotocol X.25 Adapter 149
 2963 2-port Multiprotocol X.25 Adapter 151
 2968 Ethernet 10/100 MB 153
 2969 Gigabit Ethernet - SX 154
 2970 Token-Ring High Performance Network
 Adapter 172
 2972 Auto Token-Ring LANstreamer MC 32
 Adapter 173
 2980 Ethernet High Performance LAN Adapter
 173
 2984 TURBOWAYS 100 ATM Adapter 173
 2985 Ethernet 10 MB BNC 157
 2987 Ethernet 10 MB AUI 158
 2988 ATM 155 MMF 159
 2989 TURBOWAYS 155 ATM Adapter 173
 2992 Ethernet LAN Adapter (AUI/10BaseT)
 173
 2993 Ethernet LAN Adapter 10Base2 (BNC)
 174
 2994 10/100 Ethernet Twisted Pair MC Adapter
 174
 3124 Serial port to serial port cable for draw-

- er-to-drawer connections 147
- 3125 Serial port to serial port cable for rack-to-rack connections 147
- 4008 SP Switch-8 117
- 4011 SP Switch 116
- 4012 SP Switch2 116
- 4020 SP Switch Adapter 120
- 4021 SP Switch Router Adapter 102, 121, 179
- 4022 SP Switch MX Adapter (withdrawn) 120
- 4023 SP Switch MX2 Adapter 120
- 4025 SP Switch2 Adapter 118
- 4093 Base Memory Card 66, 71
- 4110 One Pair of 128 MB DIMMs 66, 71
- 4224 Ethernet 10BaseT Transceiver 174
- 4320 One processor card with two CPUs 66, 70
- 6206 Ultra SCSI Single Ended 160
- 6207 Ultra SCSI Differential 162
- 6208 SCSI-2 F/W Single-Ended 163
- 6209 SCSI-2 F/W Differential 165
- 6212 9333 High Performance Subsystem Adapter 174
- 6214 SSA 4-Port Adapter 174
- 6215 SSA RAID 5 166
- 6216 Enhanced SSA 4-Port Adapter 174
- 6217 SSA 4-Port RAID Adapter 175
- 6219 SSA Multi-Initiator/RAID EL Adapter 175
- 6222 SSA Fast-Write Cache Module 166, 175
- 6310 ARTIC960RxD Quad Digital Trunk Adapter 167
- 8130 1.2 MB/sec Remote Asynchronous Node, 16-port, EIA-232 (US) 147
- 8131 128-port Asynchronous Controller Node Cable, 4.5 m 147
- 8132 128-port Asynchronous Controller Cable 23 cm (9 in.) 147
- 8133 RJ-45 to DB-25 Converter Cable 147
- 8134 1.2 MB/sec Remote Asynchronous Node, 16-port, EIA-232 (world trade) 147
- 8136 1.2 MB/sec Rack Mountable Remote Asynchronous Node, 16-port, EIA-232 147
- 8137 2.4 MB/sec Enhanced Remote Asynchronous Node, 16-port, EIA-232 147
- 8138 2.4 MB/sec Enhanced Remote Asynchronous Node, 16-port, RS-422 147
- 8396 RS/6000 SP System Attachment Adapter 77, 121
- 9122 Node-like Attachment 74
- 9123 Frame-like Attachment 74
- 9310 SP Switch Router Adapter Cable - 10

- Meter 179
- 9320 SP Switch Router Adapter Cable - 20 Meter 179
- feature code 2909
 - 9.1 GB Ultra SCSI disk pair 38
- feature code 2918
 - 18.2 GB Ultra SCSI disk pair 38
- feature code 3804
 - 9.1 GB Ultra SCSI 10K RPM disk pair 38
- feature code 3820
 - 36.4 GB Ultra SCSI 10K RPM disk pair 38
- file collections 182

G

- General Parallel File System 186
- Gigabit Ethernet - SX 154
- GPFS 186
- GRF 97
- Group Services 184

H

- HACMP 131
- HACWS 131
- High Availability Cluster Multi-Processing for AIX 131
- High Availability Control Workstation 131
 - limitation and restriction 133
 - software requirement 135
 - system requirement 135
- High Performance Supercomputer Systems Development Laboratory 2
- HIPPI 171
- HIPPI, one port 178
- HPF 9
- HPSSDL 2
- HSSI, two port 178

I

- IBM 9077 SP Switch Router model 04S 97
- IBM 9077 SP Switch Router model 16S 97
- IBM DB2 UDB EEE 8
- IBM Network Terminal Accelerator - 2048 Session 169
- IBM Network Terminal Accelerator - 256 Session 168
- IBM Recoverable Virtual Shared Disk 185
- IBM RS/6000 7017 Enterprise Server 73

IBM Virtual Shared Disk 185
Informix Dynamic Server AD/XP 8
installation and migration coexistence 183
Intermediate Switch Board 15
ISB 15

J

JFS 188
job flow 189
job management program 189
job priority 190
job status 190
job submitter 190
Journalled File System 188

K

Kerberos 184

L

LAPI 186, 190
linear programming 197
LoadLeveler 189
logical dependent node 13
login control 182
Lotus 1-2-3 196
Low-level Application Programming Interface 186,
190
LP 197

M

machine type
 9077 RS/6000 SP Switch Router 97
mathematical function 193
 Basic Linear Algebra Subroutines 193
 Eigensystem Analysis 193
 Fourier Transforms 193
 Linear Algebraic Equations 193
 quadratic programming 195
mathematical programming 195
 linear programming 195
 mixed integer programming 195
 network programming 195
Mathematical Programming System 196
mathematical subroutine 193
 computational chemistry 193
 computational techniques 193
 dynamic systems simulation 193

 electronic circuit design 193
 fluid dynamics analysis 193
 mathematical analysis 193
 nuclear engineering 193
 quantitative analysis 193
 reservoir modeling 193
 seismic analysis 193
 structural analysis 193
 time series analysis 193

Memory-I/O 30

Message Passing Interface 186, 190

MIP 197

mixed-Integer programming 197

model

 500 short model frame 18

 550 tall model frame 22

model class 18

Motif 190

MPI 9, 186, 190

MPS 196

MPSX/370 198

N

N+1 feature 17

national language support 191

Network File System 188

Network Installation Manager 183

Network Time Protocol 185

NFS 186, 188

NIM 183

NLS 191

NTP 185

O

Oracle 185

Oracle Enterprise Edition 8

OSL 195

P

Parallel Basic Linear Algebra Subprograms 193

Parallel Engineering and Scientific Subroutine Li-
brary 192

Parallel Environment 199

Parallel Input Output File System 187

Parallel Operating Environment 199

Parallel Optimization Subroutine Library 195

Parallel System Support Programs 181

- PBLAS 193
- PE 199
- Performance Optimization with Enhanced RISC 3
- Performance Toolbox for AIX 186
- Performance Toolbox Parallel Extensions 186, 202
- Perl 184
- PESSL 192
- physical dependent node 13
- PIOFS 187
- POE 199
- POWER 3
- POWER3 SMP High Node 35, 52
- POWER3 SMP node 48
 - 6xx bus 50
 - I/O subsystem 51
 - POWER3 microprocessor 50
 - Run Time Abstraction Software 51
 - service processor 51
 - system architecture 48
 - system firmware 51
 - system memory 50
 - system packaging 52
- POWER3-II 31
- processor node 29
- PSSP 181
- PTPE 186, 202
- PTX 186
- PVM 9, 190

Q

- QP 197
- Quadratic programming 197

R

- redundant power 17
- Requirements 91
- RJ-45 to DB-25 Converter Cable 147
- RS/6000 Cluster Technology 184
- RS/6000 SP System Attachment Adapter 77, 121
 - cable 79
 - placement restriction 78

S

- S/390 ESCON Channel Adapter 142
- S/390 ESCON Channel Emulator Adapter 171
- Scalable Electric Power Base Unit 17
- Scalable Linear Algebra Package 193

- ScaLAPACK 193
- SCSI-2 Differential External I/O Controller 170
- SCSI-2 Differential Fast/Wide Adapter/A 170
- SCSI-2 F/W Differential 165
- SCSI-2 F/W Single-Ended 163
- SCSI-2 Fast/Wide Adapter/A 170
- SCSI-2 High Performance External I/O Controller 169
- SDR 182
- SEPBU 17
- Serial port to serial port cable for drawer-to-drawer connections 147
- Serial port to serial port cable for rack-to-rack connections 147
- server consolidation 8
- Shared Memory Parallel Processing 192
- short frame 17
- Single Program Multiple Data 192
- slot 15
- Software 92
- Software Update Protocol 184
- SONET/IP OC3, one port MM fiber 177
- SONET/IP OC3, one port SM fiber 178
- SP 4
- SP frame 15
- SP Perspectives 184, 204
- SP Switch 109, 116
- SP Switch Adapter 120
- SP Switch adapter 112
- SP Switch board 110
- SP Switch chip 111
- SP Switch link 110
- SP Switch MX Adapter (withdrawn) 120
- SP Switch MX2 Adapter 120
- SP Switch network 114
- SP Switch port 111
- SP Switch Router 97
 - connection to the control workstation 105
 - connection to the SP Switch 107
 - network interface 104
 - network media card requirement 103
 - software requirement 104
 - system requirement 102
- SP Switch Router Adapter 121, 179
- SP Switch Router Adapter Cable - 10 Meter 179
- SP Switch Router Adapter Cable - 20 Meter 179
- SP Switch Router model 04S 100
- SP Switch Router model 16S 101
- SP Switch2

Hardware
 740 microprocessor 120
 MIC chip 120
 NBA chip 120
 TBIC3 chip 120
 J-TAG 116
SP Switch-8 117
SP1 4
SP2 4
SP-attached server 73
 connection to the control workstation 86
 connection to the SP Switch 88
 limitation 75
 network interface 83
 network media card requirement 79
 software requirement 81
 system requirement 76
SPS-8 117
SSA 4-Port Adapter 174
SSA 4-Port RAID Adapter 175
SSA Fast-Write Cache Module 166, 175
SSA Multi-Initiator/RAID EL Adapter 175
SSA RAID 5 166
SUP 184
System Data Repository 182
System memory 32
system partitioning 182

T

tall frame 20
Tcl 185
technical computing 9
Token Ring Auto LANstreamer 143
Token-Ring High Performance Network Adapter
172
Tool command language 185
Topology Services 184
TPC-D 8
TURBOWAYS 100 ATM Adapter 173
TURBOWAYS 155 ATM Adapter 173

U

Ultra SCSI Differential 162
Ultra SCSI Single Ended 160
user space task 190

V

V.35 cable 151
V.36/EIA 449 cable 151

W

WAN RS232 128-port 146

X

X.21 cable 151
X.25 Interface Co-Processor/2 172
X/OpenR 4.0 188
XL FORTRAN 192
XProfiler 200

IBM Redbooks review

Your feedback is valued by the Redbook authors. In particular we are interested in situations where a Redbook "made the difference" in a task or problem you encountered. Using one of the following methods, **please review the Redbook, addressing value, subject matter, structure, depth and quality as appropriate.**

- Use the online **Contact us** review redbook form found at ibm.com/redbooks
- Fax this form to: USA International Access Code + 1 914 432 8264
- Send your comments in an Internet note to redbook@us.ibm.com

Document Number	SG24-5596-01
Redbook Title	RS/6000 SP Systems Handbook
Review	
What other subjects would you like to see IBM Redbooks address?	
Please rate your overall satisfaction:	<input type="radio"/> Very Good <input type="radio"/> Good <input type="radio"/> Average <input type="radio"/> Poor
Please identify yourself as belonging to one of the following groups:	<input type="radio"/> Customer <input type="radio"/> Business Partner <input type="radio"/> Solution Developer <input type="radio"/> IBM, Lotus or Tivoli Employee <input type="radio"/> None of the above
Your email address: The data you provide here may be used to provide you with information from IBM or our business partners about our products, services or activities.	<input type="checkbox"/> Please do not use the information collected here for future marketing or promotional contacts or other communications beyond the scope of this transaction.
Questions about IBM's privacy policy?	The following link explains how we protect your personal information. ibm.com/privacy/yourprivacy/



RS/6000 SP Systems Handbook



Overview of the RS/6000 SP system

This IBM Redbook is a comprehensive guide dedicated to the RS/6000 SP product line. Major hardware and software offerings are introduced and their prominent functions discussed.

Information on available machine types, models, and feature codes

Topics covered include:
An overview of the RS/6000 SP system, information about available machine types, models, and feature codes, information about supported communication adapters for the RS/6000 SP system, and information about software available for the RS/6000 SP system.

In-depth discussion of Hardware Architecture

This book does not replace the latest RS/6000 SP marketing materials and tools. Rather, it is intended as an additional source of information that, together with existing resources, may be used to enhance your knowledge of IBM solutions for the UNIX marketplace using RS/6000 SP systems.

This publication is suitable for customers, sales and marketing professionals, technical support professionals, and Business Partners wishing to acquire a better understanding of RS/6000 SP products.

INTERNATIONAL TECHNICAL SUPPORT ORGANIZATION

BUILDING TECHNICAL INFORMATION BASED ON PRACTICAL EXPERIENCE

IBM Redbooks are developed by the IBM International Technical Support Organization. Experts from IBM, Customers and Partners from around the world create timely technical information based on realistic scenarios. Specific recommendations are provided to help you implement IT solutions more effectively in your environment.

For more information:
ibm.com/redbooks

SG24-5596-01

ISBN 0738419028