

IBM System z9 109

Technical Introduction

Hardware description

Software support

Key functions



Bill Ogden
Jose Fadel
Bill White

Redbooks



International Technical Support Organization

IBM System z9 109 Technical Introduction

July 2005

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

First Edition (July 2005)

This edition applies to the initial announcement of the IBM System z9 109.

© Copyright International Business Machines Corporation 2005. All rights reserved.

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

Contents

Notices	vii
Trademarks	viii
Preface	ix
The authors	ix
Become a published author	ix
Comments welcome	x
Chapter 1. Introduction	1
1.1 Evolution	2
1.2 z9-109 server highlights	3
1.3 zSeries comparisons	4
1.4 z9-109 server models and processor units	6
1.5 Upgrades	6
1.6 Considerations	7
Chapter 2. Hardware overview	9
2.1 System frames	10
2.2 Books	10
2.3 MCM	12
2.4 Processor units	13
2.4.1 PU characterizations	14
2.5 Memory	14
2.6 I/O interfaces	15
2.7 I/O cages and features	16
2.7.1 Physical I/O connections	18
2.7.2 Coupling connections	20
2.7.3 Cryptographic functions	20
2.8 Time functions	21
2.8.1 Sysplex Timer	21
2.8.2 Server Time Protocol (STP)	22
2.9 Hardware Storage Area	22
2.10 System control	22
2.11 HMC and SE	23
2.12 Concurrent maintenance and upgrades	24
2.13 Internal Battery Feature	24
2.14 Power and cooling	24
2.14.1 Cooling	25
Chapter 3. Key functional elements	27
3.1 Availability enhancements	28
3.2 Book removal	28
3.2.1 Book replacement	29
3.3 Enhanced driver maintenance	29
3.4 Redundant I/O Interconnect	30
3.5 Logical channel subsystems	30
3.5.1 Spanned channels	32
3.5.2 Channel definitions in the IOCP statement	33
3.6 Multiple subchannel sets	34

3.6.1 IOCP example	36
3.7 OSA for NCP	37
3.8 GARP VLAN Registration Protocol (GVRP)	38
3.9 The MIDAW facility	39
3.10 N_Port ID virtualization	41
3.11 New and changed instructions	42
3.12 Branch detection	42
3.13 Performance factors	42
3.14 Program-directed re-IPL	44
3.15 Open exchanges for FCP and FICON	44
3.16 CHPID mapping	44
3.17 LPARs	45
3.18 Hardware Management Console and Support Element	45
Chapter 4. Software support	51
4.1 Operating system support	52
4.2 60 logical partitions	53
4.3 Single system image	53
4.4 63.75K subchannels	54
4.5 Multiple subchannel sets (MSS)	54
4.6 MIDAW facility	55
4.7 N_Port ID virtualization	55
4.8 FICON link incident report	55
4.9 Program-directed re-IPL	56
4.10 OSA-Express2 1000BASE-T Ethernet	56
4.11 OSA-Express2 GARP VLAN Registration Protocol (GVRP)	56
4.12 OSA-Express2 OSN support	57
4.13 HiperSockets IPv6	57
4.14 Crypto Express2	58
4.14.1 PCI-X support	58
4.14.2 CPACF enhancements	58
4.15 Enhanced performance assists for z/VM Linux guests	59
4.16 Support summary	59
Chapter 5. General considerations	61
5.1 Hardware Management Console and Support Element	62
5.1.1 Support elements	63
5.1.2 HMC and SE connectivity	63
5.1.3 Planning recommendations	65
5.2 Cabling services and cable planning	65
5.3 I/O channel considerations	67
5.3.1 Parallel channel migration	67
5.3.2 ESCON channel planning	68
5.3.3 FICON Express2 planning	68
5.3.4 OSA-Express2 planning	69
5.3.5 Coupling link connectivity	69
5.4 I/O configuration definition and management	70
5.5 Capacity upgrades	71
5.5.1 Capacity Upgrade on Demand (CUoD)	72
5.5.2 Customer Initiated Upgrade (CIU)	72
5.5.3 On/Off Capacity on Demand (On/Off CoD)	72
5.5.4 Capacity Backup (CBU)	73
5.6 Logical planning	74

5.6.1 LPAR considerations	74
5.7 Coupling Facility planning	76
5.7.1 Sysplex Timer ETR Network ID	76
5.8 Operating system migration	77
Appendix A. Frequently asked questions	79
Appendix B. Instruction set changes	81
Appendix C. Channel options	83
Related publications	85
IBM Redbooks	85
Other publications	85
Online resources	85
How to get IBM Redbooks	85
Help from IBM	85
Index	87

Notices

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

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

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

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

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

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

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

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

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

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

COPYRIGHT LICENSE:

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

Trademarks

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

@server®	ECKD™	PR/SM™
@server®	ESCON®	Redbooks™
Redbooks (logo)  ™	FICON®	Resource Link™
eServer™	HiperSockets™	S/360™
z/Architecture™	IBM®	S/390®
z/OS®	MVS™	Sysplex Timer®
z/VM®	OS/2®	System z9™
z/VSE™	Parallel Sysplex®	VSE/ESA™
zSeries®	Power PC®	VTAM®

The following terms are trademarks of other companies:

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

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

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

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

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

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

Preface

This IBM® Redbook introduces the IBM System z9™ 109 (z9-109), which offers a continuation of the IBM scalable mainframe servers. The z9-109 server is based on z/Architecture™ and provides extensions to the architecture. In addition to increased performance and expansion options, particular emphasis has been placed on the integration of the z9-109 into a complete system and software environment.

This document provides basic information about new hardware functions and features, associated software support, and migration considerations. It is intended for systems engineers, hardware planners, and anyone else needing to understand the new z9-109 system elements and plan for their usage. This document is not intended as an introduction to mainframes. Readers are expected to be generally familiar with current zSeries® technology and terminology.

The authors

This redbook was produced by a team working at the International Technical Support Organization, Poughkeepsie Center.

Bill Ogden is a retired IBM Senior Technical Staff Member at the International Technical Support Organization, Poughkeepsie, who has worked since 1966 with the S/360™ and all its descendants.

Jose Fadel is a zSeries System Architect from Brazil. He has 28 years of experience with IBM Large Systems. During most of his 15 years with IBM, he has worked as a pre-sales technical support specialist for zSeries. His areas of expertise include zSeries hardware and software.

Bill White is a Project Leader and Senior Networking Specialist at the International Technical Support Organization, Poughkeepsie.

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

Jason Boxer
IBM Systems & Technology Group, Operations TSM Lead, IBM Poughkeepsie

Charles Shapley
IBM Systems & Technology Group, Configuration Plans & Design, IBM Poughkeepsie

Become a published author

Join us for a two- to six-week residency program! Help write an IBM Redbook dealing with specific products or solutions, while getting hands-on experience with leading-edge technologies. You'll team with IBM technical professionals, Business Partners and/or customers.

Your efforts will help increase product acceptance and customer satisfaction. As a bonus, you'll develop a network of contacts in IBM development labs, and increase your productivity and marketability.

Find out more about the residency program, browse the residency index, and apply online at:

ibm.com/redbooks/residencies.html

Comments welcome

Your comments are important to us!

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

- ▶ Use the online **Contact us** review redbook form found at:

ibm.com/redbooks

- ▶ Send your comments in an email to:

redbook@us.ibm.com

- ▶ Mail your comments to:

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



Introduction

The IBM System z9 109 (also known as the *z9-109*) is the next step in the evolution of the mainframe family. While not part of the zSeries, it builds on the z/Architecture and the system structure introduced with the z990. The physical appearance of the z9-109 server and z990 servers is very similar. However, in addition to extending zSeries technology, the z9-109 server delivers enhancements in the areas of performance, scalability, availability, security and virtualization. The z9-109 system has been designed using a holistic approach that includes the latest operating system, middleware, storage, and networking technologies. The synergies gained through this collaborative design and development enables the elements to support each other while helping to delivering additional value to your business.

Today, many businesses are starting their transformation to become on demand businesses, enabling them to sense and respond to change on the fly and with flexibility, to manage risk while lowering costs, and to gain greater control and transparency with less complicated management. Businesses are realizing their goals, by addressing complexity in the infrastructure and by creating a flexible and resilient on demand operating environment.

The scalability of the z9-109 system, along with its further advances in availability and security, can bring these on demand benefits to help align business processes and priorities. Its multi-book design provides enhanced availability and a great degree of flexibility to tailor the server to your precise needs, while the wide choice of *specialty processors* enables key workloads to be deployed at lower cost of ownership.

The z9-109 holistic system approach is designed to deliver high levels of utilization, to automatically allocate resources where and when they are needed according to business priorities, and to adapt to change quickly and reliably. This makes the z9-109 *the* data center for the on demand business world.

1.1 Evolution

The z9-109 server continues the evolution of IBM mainframe servers, with extensions in several directions. This evolution observes the key factors of mainframe progress:

- ▶ *Compatible evolution.* IBM recognizes the very large customer investment in existing data processing applications and associated business processes. As has been the case since S/360 was introduced, new functions and features have been designed to accommodate existing applications whenever possible.
- ▶ *Resolving constraints.* Evolving hardware capabilities and increasing application complexity and size continue to push the limits of existing systems. Every new generation of mainframe servers reduces various constraints that may have inhibited application growth on earlier systems.

Providing such compatibility can require considerable design ingenuity to provide system growth while retaining compatibility with earlier systems. Examples of evolution include:

- ▶ Single processor systems evolved to multiple processor systems. *The z9-109 server has continued this evolution by providing a modular multi-book design that supports one to four books and up to 54 processor units (customer-usable PUs) per server.*
- ▶ Real memory addressing evolved to virtual memory systems. *The z9-109 server expands this function by providing a more sophisticated Translation Lookaside Buffer (TLB) design that is in effect, 16-way associative.*
- ▶ Memory addressing (real and virtual) evolved from 24 bits to 31 bits to 64 bits for addressing. *The z9-109 server allows for full 64-bit real and virtual storage support, and any logical partition can be defined for 31-bit or 64-bit addressability.*
- ▶ Real memory size has grown with every IBM mainframe generation. *The z9-109 server continues this growth by providing up to 512GB of system memory.*
- ▶ A single system image has evolved to logical partitions providing multiple system images. *The z9-109 server continues this evolution by providing up to 60 logical partitions.*
- ▶ Processor speeds have grown with every generation. There is no simple metric to characterize processor performance in today's systems because many factors interact in very complex ways. *The z9-109 server clock speed has increased to 1.65 GHz, providing a significant performance improvement over earlier zSeries servers.*
- ▶ I/O channels evolved from a maximum of 16 channels in S/360 machines to a much larger number of channels. *The z9-109 server continues such growth by increasing the number of STIs (up to 16 per book) and allowing more high-speed channel features to be used in the I/O cages. Each STI has a bandwidth up to 2.7 GB/s.*
- ▶ The number of I/O devices in a system was initially limited by the number of channels, the number of control units on each channel, and the number of devices on each channel. The addressing structure also provided a limitation. The fixed three-byte addresses (one byte each for channel, control unit, and device) of early systems evolved into four-byte device numbers allowing up to almost 64K device addresses. *The z9-109 server continues this growth by providing Multiple Subchannel Sets (MSS), allowing up to almost 128K device addresses.*
- ▶ Channel performance has grown from parallel channels to ESCON® channels to FICON® channels. *The z9-109 server continues such growth by providing a significantly higher-performance option for channel programming.*
- ▶ Server workloads have been partly offloaded into segregated processors such as SAPs, ICFs, IFLs, and zAAPs. *The z9-109 server enhances the management by providing separate pools for PR/SM™ handling of shared ICFs, IFPs, and zAAPs.*

- ▶ Basic “real” systems evolved into virtual systems and this evolution has extended to systems, processors, memory, I/O devices, LAN interfaces, and so forth. *The z9-109 server continues this direction with new instructions that improve the performance of virtual machine QDIO operations. This is done by creating a passthrough architecture designed to reduce host programming overhead, avoiding the stopping of guest processing when adapter interruptions are present.*
- ▶ Recent mainframe generations have extended the instruction set provided to include instructions more compatible with other platforms (such as binary floating point), instructions to better implement popular languages (such as the string-handling instructions for C/C++), instructions to improve register usage (such as the relative and immediate instructions, and the long-displacement instructions), and so forth. *The z9-109 server continues this expansion with new and changed instructions.*
- ▶ Cryptographic hardware assistance has been available in many forms on earlier systems, and with much more emphasis in more recent servers. *The z9-109 server continues the evolution of cryptographic hardware processing by extended the functions of the basic cryptographic instructions and by consolidating the options (secure coprocessor and accelerator) in a single feature. The two options can be individually defined to the feature.*
- ▶ Transparent hardware recovery has been a keystone in mainframe design and has evolved in many directions. *The z9-109 server continues this evolution by extending such transparent recovery functions to include the paths from I/O cages to system memory.*
- ▶ Concurrent maintenance is a major design goal for modern mainframes and often involves balancing a design between replicated components and more integration onto chips and MCMs. *The z9-109 server allows for a single book, in a multi-book configuration, to be concurrently removed and reinstalled during an upgrade or repair.*
- ▶ Clustering of mainframes evolved from simple shared DASD to basic sysplex functions and then to Parallel Sysplex® operation. *The z9-109 server continues this evolution by better integrating the rigorous timer coordination¹ needed in a Parallel Sysplex.*

1.2 z9-109 server highlights

As mentioned, the z9-109 server is the next step after the zSeries. It uses the z/Architecture and instruction set (with some extensions) of the z900 and z990 servers. (This architecture, formerly known as ESAME Architecture, is commonly known as *64-bit* architecture, although it provides much more than 64-bit capability.)

The z9-109 server is offered as five hardware models and has a machine type of 2094. The model number indicates the maximum number of processor units (PUs) potentially orderable, and not the actual number that has been ordered.

A z9-109 server has two frames that are the same size as the z990 frames (see Figure 1-1 on page 4). They are air-cooled and refrigeration-cooled. The frames use three-phase power, and an Internal Battery Feature (IBF) is also provided.

Also included in the frames are:

- ▶ Up to four books, containing PUs, memory, and STI ports
- ▶ Two ThinkPads that are used as Support Elements
- ▶ Up to three I/O cages with 28 I/O slots each

Some I/O cards from a z900 or z990 can be moved forward to the z9-109 server.

¹ This function is called Server Time Protocol (STP) and is being previewed for the z890, z990, and z9-109 servers.

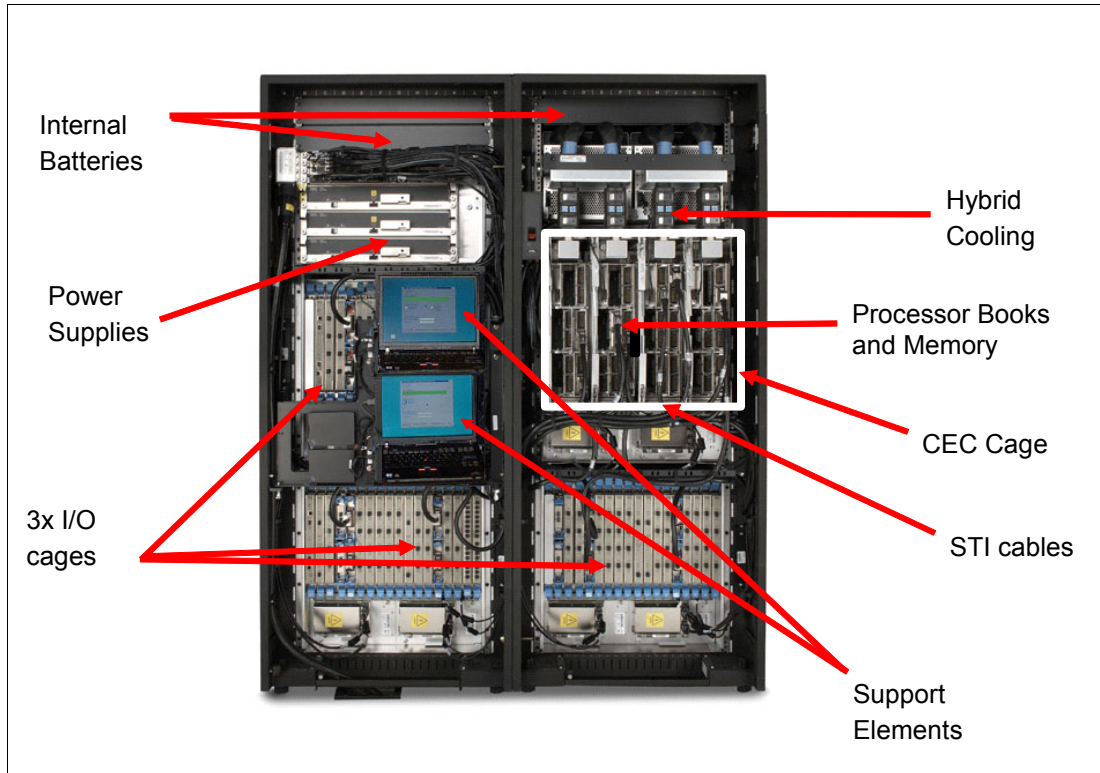


Figure 1-1 z9-109 server - under the covers - front view

A more extensive description of the frames and features is provided in Chapter 2, “Hardware overview” on page 9.

1.3 zSeries comparisons

Table 1-1 provides a brief comparison of zSeries servers to the z9-109 server.

Table 1-1 Comparison of zSeries servers

	z900	z800	z990	z890	z9-109
Machine type	2064	2066	2084	2086	2094
Maximum number of PUs	20	5	48	5	64
Max characterized PUs	16	4	32	4	54
Standard SAPs	3	1	2/book	1	2/book
Spare PUs	1	unused PUs	2/book	unused PUs	2/server
Max memory	64 GB	32 GB	64 GB/book	32 GB	128 GB /book
Cycle time	1.3/1.09 ns	1.6 ns	.83 ns	1.0 ns	.58 ns
Maximum LPARs	15	15	30	30	60
Maximum I/O cages	3	1	3	1	3

	z900	z800	z990	z890	z9-109
L1 cache (processor)	256K/256K	256K/256K	256K/256K	256K/256K	256K/256K
L2 cache (book)	16-32 MB	8 MB	32 MB/book	32 MB	40 MB/book
Logical Channel subsystems	1	1	4	2	4
Subchannel sets per LCSS	1	1	1	1	2
I/O Bandwidth (GB/s) ^a	24 GB/s	6 GB	24-96 GB/s	16 GB	43-172 GB/s
Number STIs Speed each STI	24 (1.0 GB/s)	6 (1.0 GB/s)	12-48 (2.0 GB/s)	8 (2.0 GB/s)	16-64 (2.7GB/s)

a. The number shown is simply the number of STIs multiplied by the speed of the STI. It is unlikely that any practical system or application would stress these numbers. Perhaps a better title would be something like *Memory Bandwidth to I/O Interfaces*.

Tables like this can be interesting, but it is important to understand the basic details behind each of the numbers. Most of the numbers are meaningful only when used in the context of z9-109 servers and zSeries machines.

Functional comparison

Figure 1-2 summarizes key functional differences between the z9-109 server and the z990.

Processor / Memory	<ul style="list-style-type: none"> Uniprocessor Performance Processor Design Models Processing Units (PUs) Memory 	<ul style="list-style-type: none"> Performance improvements over z990 SW models (301 + 332) z9-109 adds instructions, new technology, larger L2 cache z9-109 has 5 models, z990 has four models z9-109 has up to 54 PUs to configure, up to 32 on z990 z9-109 has up to 512GB, up to 256GB on z990
Virtualization	<ul style="list-style-type: none"> LPARs FCP PR/SM Dispatching 	<ul style="list-style-type: none"> z9-109 has up to 60 available LPARs, up to 30 on the z990 z9-109 supports FCP N-port virtualization z9-109 has separate PU pools for zAAPs, IFLs, and ICFs
I/O	<ul style="list-style-type: none"> HiperSockets FICON for SANs OSA-Express for LANs Total channels Internal I/O Bandwidth Bus support for I/O cards 	<ul style="list-style-type: none"> Both support 16 HiperSockets, z9-109 adds IPv6 support Up to 336 FICON Express2 channels on z9-109, up to 240 on z990 Both support 48 OSA ports, z9-109 adds OSA-Express2 1000BASE-T Both support up to 1024 channels (with 4 CSSs) z9-109 has up to 64 STIs at 2.7 GB/s, z990 has 48 STIs at 2GB/s z9-109 has redundant I/O interconnect for availability, z990 Single I/O Bus
On Demand / RAS	<ul style="list-style-type: none"> Capacity Backup Repair failed book MBA repair Change book memory cards or MCM Driver Upgrade 	<ul style="list-style-type: none"> For CPs, zAAPs, IFLs, and ICFs on z9-109, CPs only on z990 Concurrent after restart on z9-109, outage on z990 (Note 1) Concurrent on z9-109, unscheduled outage on z990 Vary the book off and continue to run on remaining books on z9-109, scheduled outage on z990 of entire machine (Note 1) Concurrent on z9-109 versus, scheduled outage on z990 (Note 2)
<p>Note 1: Possible only with multiple books. May be disruptive to some workloads depending on book configuration. Note 2: Concurrent driver upgrades possible only at certain MCL levels.</p>		

Figure 1-2 z9-109 server functional comparison to z990

1.4 z9-109 server models and processor units

Five models of the z9-109 server are available (see Table 1-2). The models correspond to the maximum number of processor units (PUs) *that can be ordered* in the server. PUs are briefly described in 2.4, “Processor units” on page 13.

Table 1-2 Available models

Model	Books	Processor units ^a	I/O cages	Max channels	Max memory
S08	1	1 - 8	1 - 3	960	128 GB
S18	2	1 - 18	1 - 3	1024	256 GB
S28	3	1 - 28	1 - 3	1024	384 GB
S38	4	1 - 38	1 - 3	1024	512 GB
S54	4	1 - 54	1 - 3	1024	512 GB

a. In this description, processor units includes CPs, IFLs, ICFs, zAAPs, additional SAPs (above the minimums), and ordered-but-not-activated PUs.

Each book always includes two SAPs, which are not counted in the processor unit column of this table. Two spare PUs are always included with a system; these are not counted in the processor unit column of the table. This table provides a simplistic view of processor unit assignment; many other arrangements are available. For example, a model S18 could be ordered with only a single processor unit (not counting SAPs) in the second book.

In this table, Max PUs means the maximum number of PUs that can be configured for normal uses such as CPs, IFLs, ICFs, zAAPs, additional SAPs, or various unconfigured PU options. This table reflects PU usage for ordering purposes. Once installed, the allocation of the two spare PUs may be in any installed book.

Note that the model number is not related to the number of processor units (CPs, IFLs, ICFs, zAAPs) that are characterized. The *model numbers* are S08, S18, S28, S38, S54. The STSI instruction may be used to obtain a number related to the number of CPs enabled; such numbers are sometimes referenced as *software capacity indicators*. The numbers returned by STSI range from 700 through 754 and correspond to 0 through 54 CPs enabled. (A server might have only IFLs and/or ICFs enabled and would report zero CPs.)

1.5 Upgrades

To protect investments in zSeries technology, the following upgrades are offered with the z9-109 server models:

- ▶ Any-to-any upgrade from z900 (except model 100)
- ▶ Any-to-any upgrade from z990
- ▶ Full upgrades within the z9-109 server models

Figure 1-3 on page 7 show the upgrade path to the z9-109 server models. Note that upgrading to an S54 (planned availability November 2005) from other z9-109 models will require a planned outage.

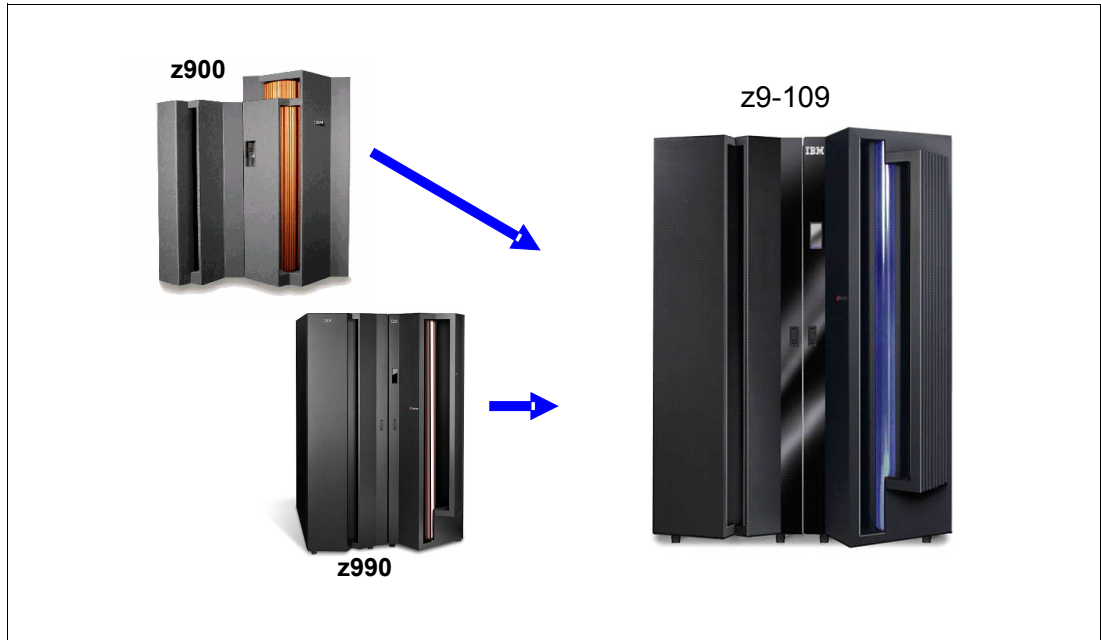


Figure 1-3 Upgrade path to a z9-109

1.6 Considerations

There are a number of other considerations involved when migrating from a z900 or earlier system to z9-109. These include the following:

- ▶ Parallel (“bus and tag”) channels are not available. Converters (which convert an ESCON channel to a parallel channel) may be used. These include the IBM 9034 converter and converters from Optica Technologies, Incorporated.
- ▶ Hardware cryptographic functions are implemented differently:
 - CCF cryptographic coprocessors are not available.
 - The PCICC, PCICA, and PCIXCC cryptographic features used with earlier zSeries machines are not used with the z9-109 server. They have been replaced with the Crypto Express2 feature.
- ▶ Some earlier I/O features may be carried forward into z9-109 servers, while others cannot, such as:
 - OSA-Express Token Ring feature
 - OSA-2 features (FDDI and Token Ring)
 - ICB-2 feature
 - OSA-Express 155 ATM features
 - ISC-3s may not be defined in compatibility mode (that is, CFS/CFS CHPID types)
- ▶ The SE and HMC software is completely new and is no longer based on OS/2® systems. While the general concepts and functions remain approximately the same, the displays and detailed operational characteristics are somewhat different. Some training and time for familiarity may be needed for operators familiar with previous HMC techniques.

Migration

Migration from a z990, which supports multiple logical channel subsystems should be straightforward. However, if multiple subchannel sets are to be implemented, this may require additional planning. Migration from a z900 involves extending existing HCD/IOCP/IOCDS

definitions to include the physical channel identifiers (PCHIDs) associated with logical channel subsystems. This is probably the most complex component of a migration to a z9-109 server.

For more details, refer to Chapter 5, “General considerations” on page 61.

Hardware overview

The z9-109 server represents an evolution beyond the zSeries servers, with the z990 as a starting point for many elements. This chapter describes key hardware elements and compares them with zSeries servers where relevant. Figure 2-1 illustrates the conceptual relationships of the major elements described in this chapter. (This figure should not be taken as anything more than a basis for discussion.)

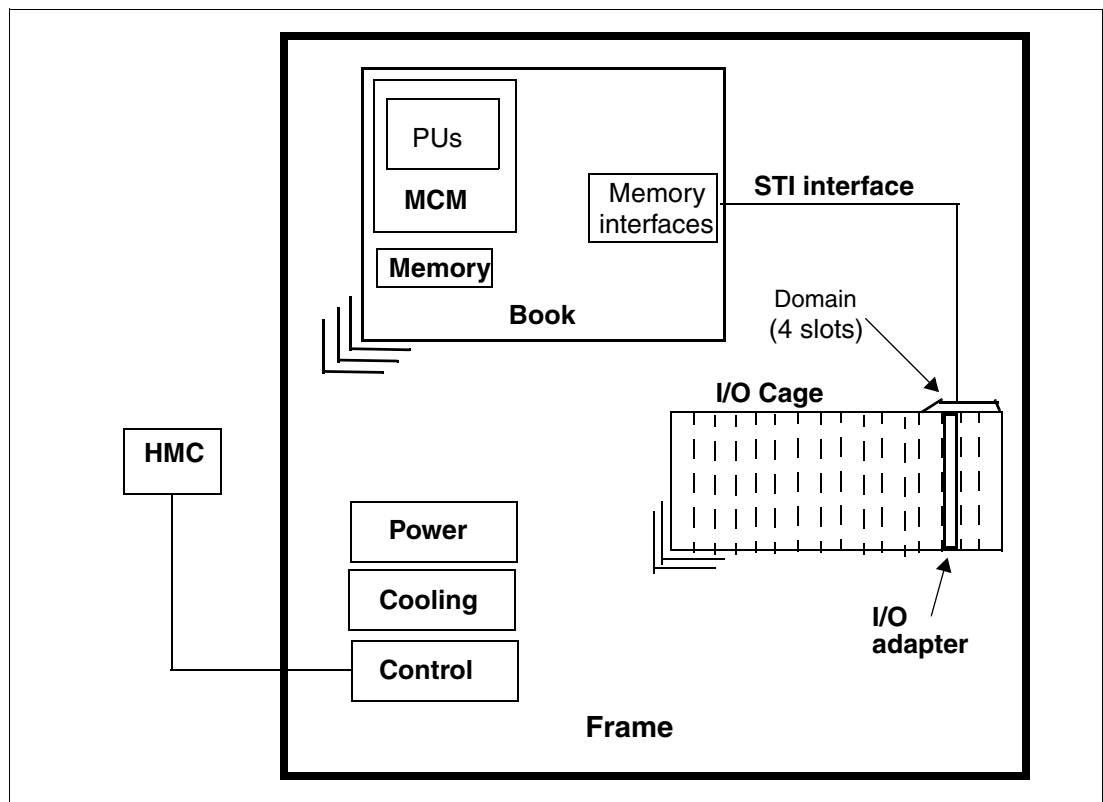


Figure 2-1 Conceptual z9-109 server component structure

2.1 System frames

The z9-109 server frames are very similar to a z990, as illustrated in Figure 2-2.

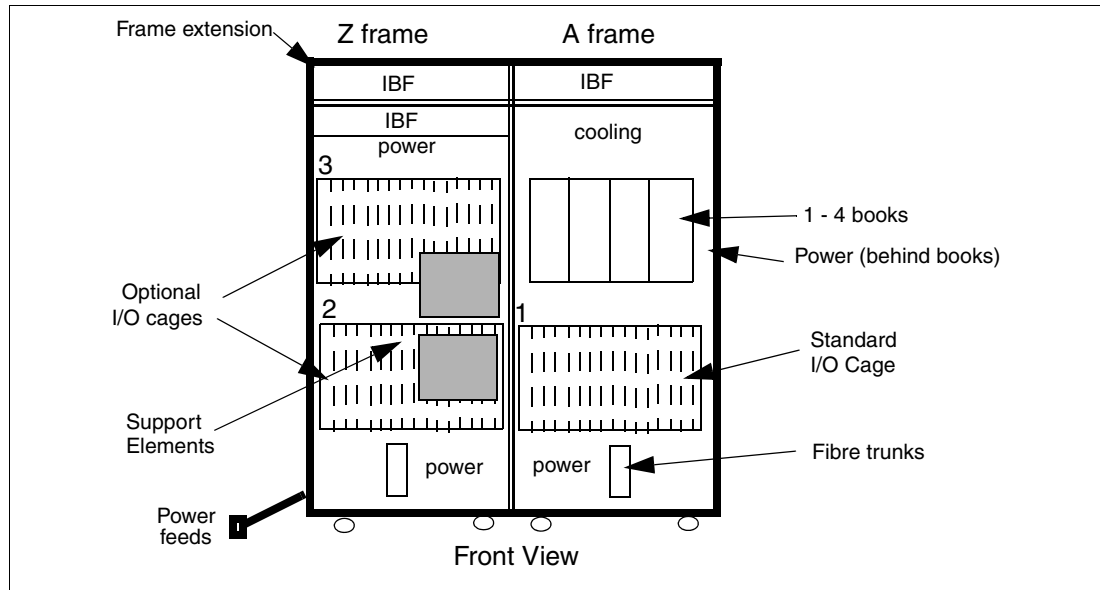


Figure 2-2 System frames

The footprint, service clearances, and channel cabling are the same as with a z990. There are always two frames, even if only one I/O cage is present. The frames are 40 EIA units tall; the two EIA units may be removed if necessary. The specific dimensions are:

- ▶ 158 cm deep (62 inches), including the covers
- ▶ 194 cm high (76 inches), including the casters
- ▶ 154 cm wide (61 inches), with both frames together
- ▶ 937kg (2060 pounds) for the A frame (without the internal battery feature)
- ▶ 1025 kg (2257 pounds) for the A frame (with the internal battery feature)
- ▶ 829 kg (1803 pounds) for the Z frame (without the internal battery feature)
- ▶ 999 kg (2197 pounds) for the Z frame (with the internal battery feature)

These approximate weights assume a full load of I/O features but do not include the covers. The weights may change slightly depending on the exact mix of I/O features.

2.2 Books

A z9-109 server has one to four *books*, depending on the model. A book contains processors (on an MCM), memory, L2 cache (also on the MCM)², ring connections to the L2 cache on other books, and memory buffer interfaces for I/O connections. The z9-109 server book design differs from the z990 design and is outlined in Figure 2-3.

² L1 cache is a part of every PU. PUs are on the MCM in a book, so L1 cache is also in every book.

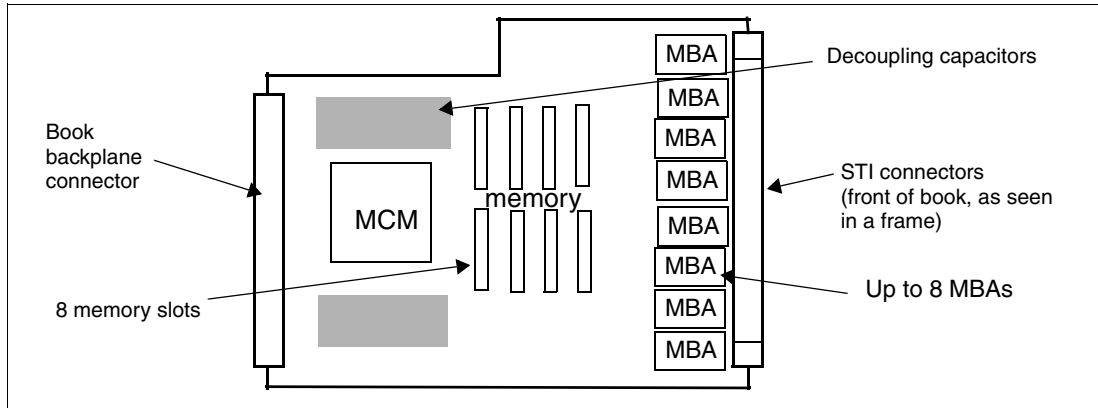


Figure 2-3 Book card layout

The book card is approximately 38 x 47 cm. It is placed in a unit that provides support for the various wiring and cooling connectors. This unit is more open than what was used in z990 and z890 servers. *Jumper books* are used for servers having fewer than four operational books, providing required wiring and cooling loops for each book slot.

L2 cache is a book-level function and has been enhanced over previous systems. Each book contains 40 MB of L2 cache memory, which is used with 20-way set associative assignments. The z9-109 server has doubled the L2 cache interleaving access from that provided with the z990 for improved performance. L2 cache is critical to the design of the z9-109 server. The L2 cache in the books is connected together via a ring design. (There are two rings, going in opposite directions. This provides both performance and redundancy.) Memory references to other books are handled through the L2 cache operation. In a sense, there is a single unified L2 cache that extends across all installed books and this extended L2 is the interface between PUs (in any book) and memory (in any book).

The number of books in a system is a direct function of the model number, as detailed in Table 2-1.

Table 2-1 Model characteristics

Model	System Books	System Spares	System SAPs	CPs, ICFs, IFLs, zAAPs, add'l SAPs ^a
S08	1	2	2	1 - 8
S18	2	2	4	1 - 18
S28	3	2	6	1 - 28
S38	4	2	8	1 - 38
S54	4	2	8	1 - 54

a. The maximum number of ICFs is 16. The maximum number of zAAPs is half the number of CPs that are characterized. The maximum number of additional SAPs is two per book.

Notice that model S54 has four books, just like model S38. The fundamental difference is that all the processor chips in the S54 are dual core, providing 16 PUs in each MCM. The other models use a mixture of single-core and dual-core chips in their MCMs. In all cases, two PUs *per book* are reserved as SAPs and two PUs *per system* are reserved as spares.

The last column of Table 2-1 indicates the number of PUs that may be characterized when the system is ordered, including those allocated for various dynamic upgrades. It would be

unusual to order a model S54, for example, that includes only one CP, but this would be possible.

Each book contains up to eight I/O *fanout* cards. Each fanout card contains an Memory Bus Adapter (MBA) module and two STI ports for a maximum of 16 STI ports per book. (The z990 was arranged differently: it had three internal MBAs and 12 fixed STI ports per book.) The increase to up to 16 STI ports per book is a significant enhancement. Also, the STI ports each operate at 2.7 GB/sec.³ (The z990 STIs operated at 2.0 GB/sec.) The conceptual fanout arrangement, from the memory subsystem to I/O adapters, is shown in Figure 2-4.

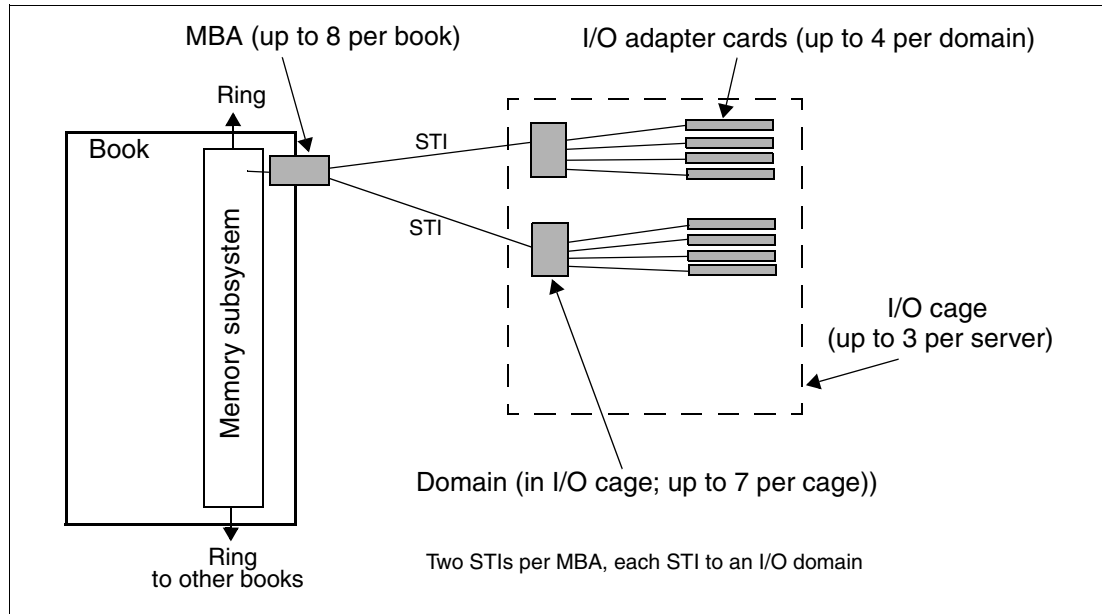


Figure 2-4 Fanout from memory to I/O adapters

IBM determines the number of MBA elements and the arrangement of I/O adapters in domains.

2.3 MCM

The MCM (Multiple Chip Module) is the heart of the server. An MCM contains all the processor chips and L2 cache of a book. The MCM itself is a 104-layer glass ceramic carrier that is the product of very sophisticated materials engineering. The layout of chips on an MCM is shown in Figure 2-5; it is unchanged from the general z990 layout, although the technology in the chips is new.

³ Except when the STI is used for ICB4 or ICB3 coupling, in which case it operates at 2.0 GB/second or 1 GB/s, respectively.

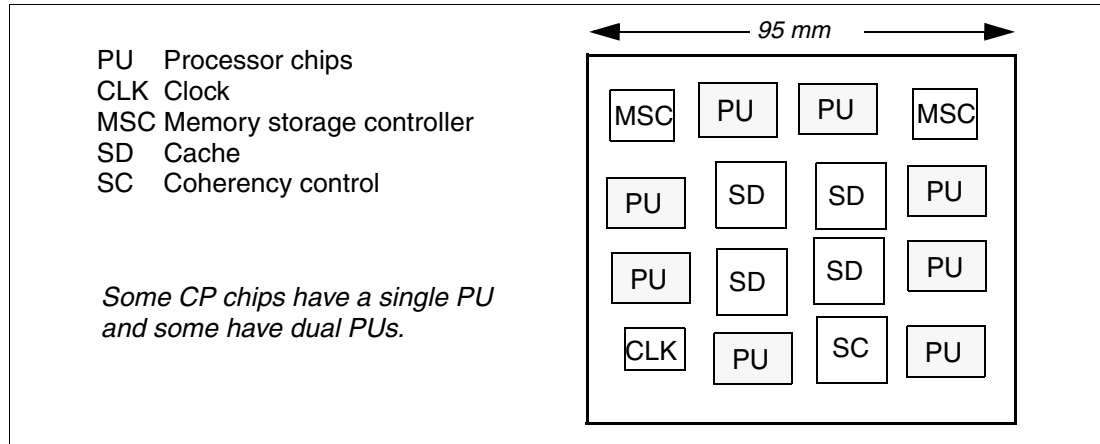


Figure 2-5 MCM layout

The MCM contains 8 processor chips. These are a mixture of single-core and dual-core processor chips providing a total of 12 PUs, except in model S54 systems where only dual-core chips are used. Two PUs are always used as SAPs. The remaining PUs may be characterized in various ways, depending on the features ordered with the server, or may remain uncharacterized. Two PUs in the server are reserved as spares. (Earlier servers reserved two spare PUs *on each book*. This has proven unnecessary, thus two spares *per system* are reserved in the z9-109 server.)

An MCM has 5184 connection pins for signals, power, and grounds. MCMs in all books contain the same functions.

2.4 Processor units

Processor chips may have one or two processor *cores* on each chip. Each core is a z9-109 processor unit (PU). Each PU has an internal mirror function whereby most instructions are, in effect, executed twice for RAS purposes. This internal mirror function in each PU should not be confused with dual-core chips, which contain two complete PUs (each with the mirror function). The following discussion is for each PU and is independent of whether it exists on a single-core or dual-core chip.

Each PU is a superscalar processor with these characteristics:

- ▶ The basic cycle time is approximately 580 picoseconds.
- ▶ Up to two instructions may be decoded per cycle.
- ▶ Up to three instructions may be executed (finished) per cycle.
- ▶ Instructions are completed in the order in which they appeared in the instruction stream.
- ▶ Memory accesses might not be in the same instruction order.
- ▶ Most instructions flow through a pipeline with different numbers of steps for various types of instructions. A number of instructions may be in progress at any instant, subject to the maximum number of decodes and completions per cycle.

Each PU has two L1 caches. There is a 256 KB cache for instructions and a different 256 KB cache for data. Each L1 cache is four-way set associative, and has its own Translation Lookaside Buffer (TLB) of 512 entries (also four-way set associative). There is also a secondary TLB with 512 entries at the first virtual memory segment (or region) level and 4096 entries at the page level. The secondary TLB is, in effect, 16-way set associative.

A branch history table, with 8K entries, is four-way set associative and enhances the address translation function. The general address translation functions have been moved to a separate engine (driven by “pico code”) for more flexibility in handling address translation.

There are new and/or changed instructions introduced with the z9-109 server, in the following categories:

- ▶ Several instructions for unnormalized HFP multiply and multiply/add
- ▶ Several instructions providing extended 32-bit immediate functions
- ▶ More translate instructions for Unicode support
- ▶ Extensions to the cryptographic instructions
- ▶ New system control instructions

2.4.1 PU characterizations

Processor units are ordered in single increments. A book always has a fixed number of processors, but they are activated only when ordered. (The two fixed SAPs in each book are always activated.) The internal server functions, based on the configuration ordered from IBM, *characterize* processors into various types during a power-on reset (POR) function. A processor that is not characterized cannot be used. A PU may be characterized as follows:

- ▶ CP - These are the “normal” z9-109 processors for use with an operating system and user applications. They may be used for any z9-109 operating system, including Linux®.
- ▶ ICF - These are Coupling Facility processors and are dedicated for this purpose. An ICF does not “count” when establishing the capacity of the server (often based on the number of CPs) used for software pricing.
- ▶ IFL - These are used for Linux processing and for limited z/VM® processing. (The intention is that z/VM is used to host multiple Linux guests.) An IFL does not “count” when establishing the capacity of the server (often based on the number of CPs) used for software pricing.
- ▶ SAP - These are System Assistance Processors that may be used to increase I/O capacity. Each book contains two SAP processors by default. More SAPs may be added by characterizing a PU as an SAP.
- ▶ zAAP - These are processors that may be dispatched by z/OS® to execute JVM™ functions. They may not be used for other z/OS work (or by any other operating system). Like an IFL, they do not “count” when establishing the capacity of the server for the purpose of software pricing.
- ▶ Uncharacterized PUs may be ordered for future use through a number of different plans that allow temporary or permanent upgrading of the server.

The z9-109 server processor characterization functions are the same as for z990 and z890 servers. However, the ability to restructure the characterizations when the server is running in a reduced mode (due to a component failure or book maintenance activity) is more sophisticated than with earlier servers.

2.5 Memory

Each z9-109 server book contains eight slots for memory cards. Memory cards are produced in 4, 8, and 16 GB sizes. Either four or eight cards are present in a book. (This requirement is due to the fact that the memory access and RAS functions operate across four cards.)

Different books may have different amounts of memory installed. The memory cards within a book should be the same size.⁴

The largest memory possible for a book is eight cards of 16 GB each, making 128 GB of memory in the book. If this is done for four books, the total will be 512 GB of memory, which is the maximum size for a z9-109 server.

The base server is supplied with 16 GB of memory, and additional memory is ordered in 16 GB increments. IBM selects the combination of memory cards that will be used. There may be more physical memory installed than ordered. If so, a memory upgrade (within the physically installed memory) may be ordered and installed concurrently with server operation.

Storage protect keys are part of main memory with the z9-109 server. Each book contains 128K storage protect keys with 16-way set associative access.

A Flexible Memory Option (selected when ordering a server) provides extra memory that is not normally activated. Sufficient extra memory is added such that, if a book is *fenced* or removed, the server can still provide the normally active amount of memory.

2.6 I/O interfaces

Each book in a z9-109 server may have up to 16 STI interfaces. These are installed in pairs and the number provided is selected by IBM to meet the needs of the installed I/O cages, features, and ICB4/ICB3 coupling links. The configuration required for the STI connections differs from zSeries servers.

In general, an STI drives a *domain* in an I/O cage. A domain consists of four I/O feature slots. The STI links for two domains are tied together for failover purposes, as shown in Figure 2-6.

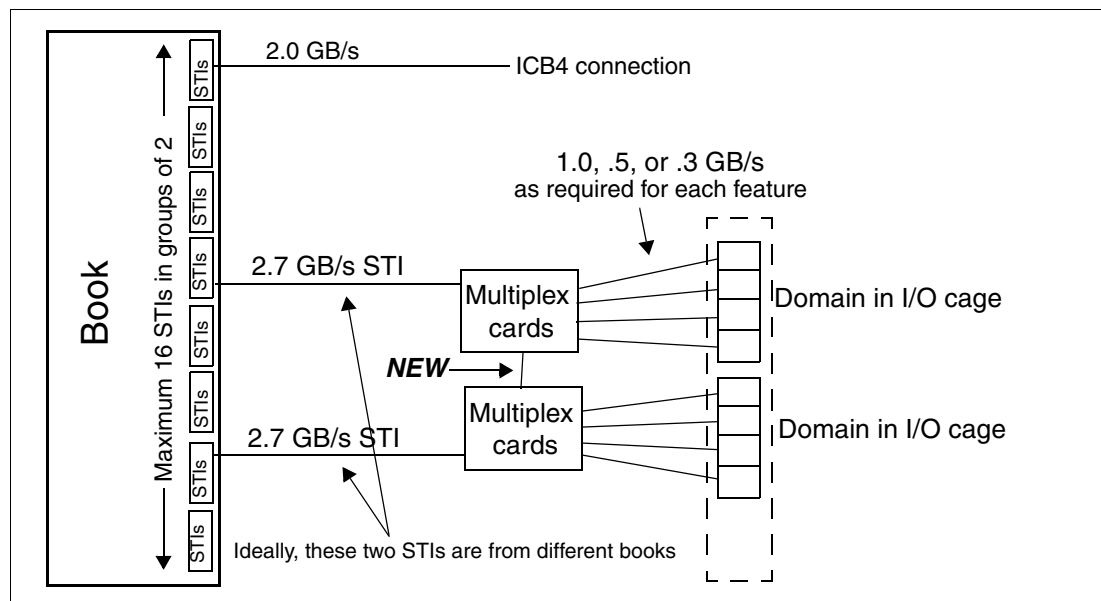


Figure 2-6 STI usage

Note the connection between the two multiplex functions in Figure 2-6. This connection provides a new redundant I/O interconnect function. If one of the STI connections fails (or is

⁴ If they are not the same size then the memory used on each card will correspond to the memory present in the smallest card.

disconnected for some reason) the other STI connection can service both domains in the I/O cage. In this case the data rate of a single z9-109 server STI (2.7 GB/s) must be shared across both domains. The redundant I/O interconnect potential is best if the two STIs are connected to different books.

An I/O cage has seven domains; a fully-populated I/O cage required seven STI connections in earlier servers. In order to have the redundant I/O interconnect function available to all domains, an eighth STI connection is required for a z9-109 server I/O cage in which all seven domains are used. This eighth STI connection is used only in case the primary STI for the last domain is disabled for some reason.

IBM determines the number of STI connections needed and each specific connection to the I/O cages based on the number and type of I/O features ordered.

2.7 I/O cages and features

The I/O cages for the z9-109 server are generally the same as for zSeries machines except that increased power is available in the cages. This allows the maximum number of certain I/O features to be increased. The layout of an I/O cage is shown in Figure 2-7.

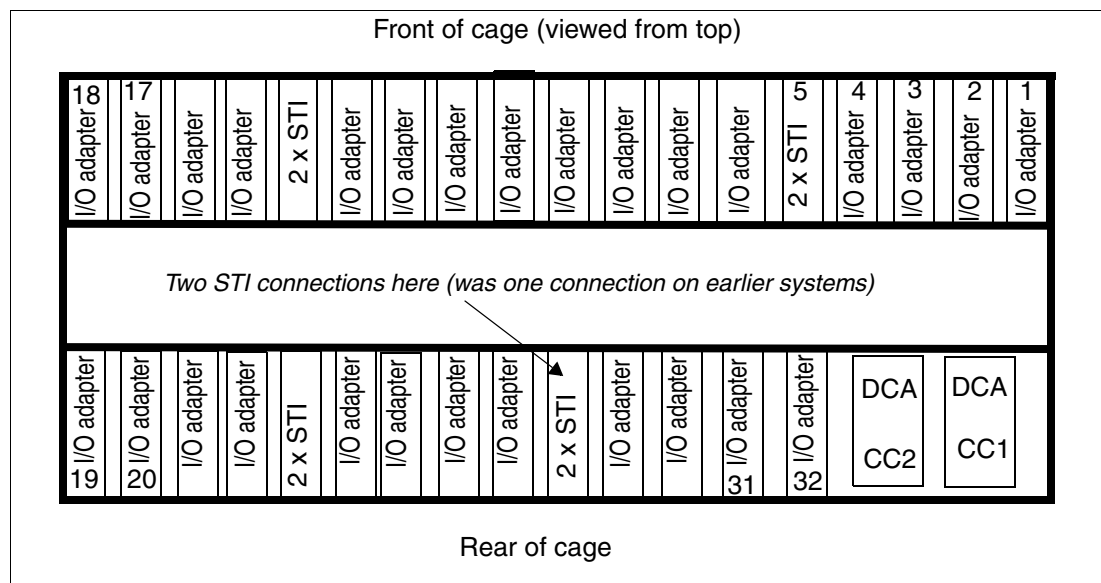


Figure 2-7 I/O cage layout

Note that slots 5, 14, 23, and 28 are used for STI connections. Assuming a fully-loaded I/O cage, there are two STIs in each of these slots. The STI connections in slot 28 serve only one domain (slots 29-32), but two STI connections are needed to provide redundant I/O interconnect.

The assignment of slots to domains for each I/O cage is as follows:

Domain	I/O slots in domain
0	1, 3, 6, 8
1	2, 4, 7, 9
2	10, 12, 15, 17
3	11, 13, 16, 18
4	19, 21, 24, 26
5	20, 22, 25, 27
6	29, 30, 31, 32

At least one major I/O feature is required with a z9-109 server. This could be an ESCON, FICON Express, FICON Express2, ICB-4, ICB-3, or ISC-3 feature. Table 2-2 shows the I/O features and interfaces used with the z9-109 server.

Table 2-2 I/O features

Feature	Max Slots	Maximum Connections	Ports/links per Feature	Purchase Increments
16-port ESCON	69 ^a	1024 ports	16 ports ^b	4 ports ^c
FICON Express2 (LX and SX versions)	84 ^d	336 ports	4 ports	1 feature
FICON Express (LX and SX versions)	60	120 ports	2 ports	1 feature
OSA-Express2	24 ^e	48 ports	2 ports ^f	1 feature
OSA-Express	24 ^e	48 ports	2 ports	1 feature
Crypto Express2	8	16 PCI-X	2 PCI-X	1 feature
ICB-4 link	N/A ^g	16 links ^h		1 link
ICB-3 link ⁱ	8	16 links	2 links	1 link
ISC-3	12	48 links	4 links	1 link
HiperSockets™	N/A	16	N/A	N/A
IC channel (CF link)	N/A	32	N/A	N/A

- a. A maximum of 64 ESCON features is possible on a z9-109 server model S08.
- b. Each ESCON feature has 16 channels, of which a maximum of 15 may be activated.
- c. The purchased number of channels are activated via Licensed Machine Code Configuration Control (LCC). Channels are activated equally across all installed ESCON features.
- d. A maximum of 64 FICON features is possible on a z9-109 server model S08.
- e. The total of OSA-Express2 and OSA-Express features is limited to 24 per server.
- f. The 10 Gb Ethernet feature has one port per feature.
- g. ICB-4 links do not involve an I/O cage.
- h. There is a maximum of 64 CF links (ICs, ICB-3s, ICB-4s, and active ISC3s).
- i. STI-3 cards (in an I/O cage) are used with ICB-3 links.

Note: The z9-109 will be the last server to support Parallel Sysplex connectivity using the ICB-3 links.

FICON Express and FICON Express2 features may be carried forward when migrating from zSeries machines. They operate at 1 or 2 Gb/second, depending on the link and the control unit involved.

OSA-Express and OSA-Express2 features may serve various functions (CHPID types), shown in Table 2-3.

Table 2-3 OSA-Express features

Feature	Media	Ports	Supported CHPID types
OSA-Express2 10 GbE LR	fiber	1	OSD
OSA-Express2 GbE LX, SX	fiber	2	OSD, OSN
OSA-Express2 1000BaseT Ethernet	copper	2	OSD, OSE, OSC, OSN

Feature	Media	Ports	Supported CHPID types
OSA-Express GbE LX, SX	fiber	2	OSD
OSA-Express 1000BaseT Ethernet	copper	2	OSD, OSE, OSC

OSA-Express and OSA-Express2 features may be carried forward when migrating from zSeries machines. The external functions are the same, but the OSA-Express2 features have a more powerful internal engine. Notice that no token ring support is provided with the z9-109 server.

The CHPID types listed in this table have the following meanings:

- ▶ OSD indicates QDIO usage by the operating system.
- ▶ OSE indicates non-QDIO usage (that is, CCW-based), by the operating system.
- ▶ OSC indicates TN3270E functions that provide support for operating system consoles and 3270 user sessions (for TSO, for example).
- ▶ OSN indicates a channel used for a link to the Communication Controller for Linux (CCL) on z9-109. This function is not available on zSeries servers.

Note that LAN adapters with fiber media are used only through QDIO, while adapters with copper media may be used with either CCW or QDIO programming. The basic characteristics are as follows:

- ▶ The adapter includes a single port.
- ▶ It uses CHPID type OSD only. (This means that only QDIO operation is allowed.)
- ▶ It uses 9-micron single-mode fiber with an SC Duplex connector.

The OSA-Express2 adapters include the following enhancements:

- ▶ Each port provides up to 640 TCP/IP stacks, improving the virtualization reach of the port.
- ▶ A *large send* function allows TCP/IP segmentation to be offloaded into the adapter.
- ▶ Layer 2 support exists for protocol-independent packet forwarding.
- ▶ LIC updates may be performed concurrently with server operation.

The following are I/O configuration details for the z9-109 server:

- ▶ A maximum of 64 coupling links can be used.
- ▶ Token ring features are not available.
- ▶ No ICB-2 links are available.
- ▶ FICON bridge support is not available with FICON Express2 features. It is available with the FICON Express LX feature when upgrading from zSeries machines.
- ▶ OSA-Express and OSA-Express2 features must not exceed a maximum of 24 for the server.

The OSA/SF software function is not required for OSA-Express or OSA-Express2 when operating in QDIO mode. However, we recommend its use for status and configuration information.

2.7.1 Physical I/O connections

The concept of logical channel subsystems (LCSSs) was introduced with z990 servers. This function is unchanged with the z9-109 server.

Logical channel subsystems add another element to the addressing of channels (CHPIDs). There is now a physical channel identifier (PCHID) that is directly related to the physical location of the channel feature in an I/O cage. IOCDS parameters then relate PCHID numbers to CHPID numbers. CHPID numbers are arbitrarily assigned when building an IOCDS.

Each possible “physical” CHPID port/connector has a PCHID number that is fixed by the physical location of the feature and connector in an I/O cage. An IOCDS associates a CHPID number to a PCHID number. A feature in an I/O cage slot can provide a maximum of 16 CHPIDs. (The ESCON feature does this, although only 15 can be used at one time.) There are a maximum of three I/O cages.⁵ Each cage has a maximum of 28 slots. PCHID numbers are assigned as follows:

I/O cage slot	-----PCHID numbers-----		
	I/O Cage 1	I/O Cage 2	I/O Cage 3
1	100 - 10F	300 - 30F	500 - 50F
2	110 - 11F	310 - 31F	510 - 51F
3	120 - 12F	320 - 32F	520 - 52F
4	130 - 13F	330 - 33F	530 - 53F
5			
6	140 - 14F	340 - 34F	540 - 54F
7	150 - 15F	350 - 35F	550 - 55F
8	160 - 16F	360 - 36F	560 - 56F
9	170 - 17F	370 - 37F	570 - 57F
10	180 - 18F	380 - 38F	580 - 58F
11	190 - 19F	390 - 39F	590 - 59F
12	1A0 - 1AF	3A0 - 3AF	5A0 - 5AF
13	1B0 - 1BF	3B0 - 3BF	5B0 - 5BF
14			
15	1C0 - 1CF	3C0 - 3CF	5C0 - 5CF
16	1D0 - 1DF	3D0 - 3DF	5D0 - 5DF
17	1E0 - 1EF	3E0 - 3EF	5E0 - 5EF
18	1F0 - 1FF	3F0 - 3FF	5F0 - 5FF
19	200 - 20F	400 - 40F	600 - 60F
20	210 - 21F	410 - 41F	610 - 61F
21	220 - 22F	420 - 42F	620 - 62F
22	230 - 23F	430 - 43F	630 - 63F
23			
24	240 - 24F	440 - 44F	640 - 64F
25	250 - 25F	450 - 45F	650 - 65F
26	260 - 26F	460 - 46F	660 - 66F
27	270 - 27F	470 - 47F	670 - 67F
28			
29	280 - 28F	480 - 48F	680 - 68F
30	290 - 29F	490 - 49F	690 - 69F
31	2A0 - 2AF	4A0 - 4AF	6A0 - 6AF
32	2B0 - 2BF	4B0 - 4BF	6B0 - 6BF

PCHID numbers used for direct book connections: 000-0FF

Note that slot numbers 5, 14, 23, and 28 are used for STI connectors.

Consider an ESCON feature in slot 15 of I/O cage 1. PCHID numbers 1C0 - 1CF are reserved for this slot. The first ESCON connector on the feature (whether or not that particular port is enabled or used) is PCHID 1C0, the second connector is 1C1, and so forth. The PCHID numbers are fixed. The CHPID numbers are not fixed and can be arbitrarily

⁵ The numbering of the I/O cages is shown in Figure 2-7 on page 16.

assigned when an IOCDS is constructed. For example, an installation might decide to make PCHID 1C0 (the first ESCON channel in our example) CHPID 52 in LCSS 1.

ICB4 connections to books

Direct book connections (to STI ports) are used for ICB-4 channels. There are up to 16 STI ports on a book, and a PCHID number is assigned to each one, as follows:

book	PCHIDs
0	010 - 01F
1	020 - 02F
2	030 - 03F
3	000 - 00F

These PCHIDs and connections do not involve I/O cages.

2.7.2 Coupling connections

The following coupling connections are supported:

Table 2-4 Coupling connections

From/To	G5/G6	z900	z990	z9-109
G5/G6	ISC2 ICB2 IC			
z900, z800	ISC comp ICB2	ISC3 peer ISC3 comp ICB2 ^a ICB3 IC		
z990, z890	ISC3 comp ICB2	ISC3 peer ISC3 comp ICB3	ISC3 peer iISC3 comp ICB3 ICB4 IC	
z9-109		ISC3-peer ICB3	ISC3 peer ICB3 ICB4	ISC3 peer ICB3 ^b ICB4

a. Not supported on z800 and not recommended between z900s because ICB3 offers better performance.

b. The z9-109 server is the last system that will support ICB3 connections.

ICB3 usage is not recommended in situations where ICB4 could be used. ICB4 offers better performance.

2.7.3 Cryptographic functions

Every processor in a z9-109 server has the CP Assist for Cryptographic Function (CPACF) instructions introduced with the z990 and these instructions have been extended with the following additional functions:

- ▶ Advanced Encryption Standard operations (AES)
- ▶ Secure Hash Algorithm - 256 (SHA-256)
- ▶ Pseudo Random Number Generation

A new optional cryptographic feature is available. It replaces the PCICC, PCICA, and PCIXCC cryptographic features available on zSeries servers. Only the Crypto Express2 feature may be used with the z9-109 server.

The Crypto Express2 feature contains two PCI-X adapters (not to be confused with the two optional internal CCF coprocessors available on several earlier machines). Each Crypto Express2 adapter can be configured as one of the following:

- ▶ *Secure coprocessor*, with functions similar to the older PCICC or PCIXCC cryptographic feature, including the secure key functions. This mode is intended for Federal Information Processing Standard (FIPS) 140-2 Level 4 certification.
- ▶ *Accelerator*, with functions similar to those provided by the older PCICA feature. This mode is usually intended for public key handshakes, as used with Web processing.

A new Trusted Key Entry (TKE) workstation is available that supports the Crypto Express2 feature and a new Graphical User Interface (GUI). This new workstation provides level 5.0 of the TKE implementation.

2.8 Time functions

Time functions are used to provide an accurate time-of-day value and to ensure that the time-of-day value is properly coordinated among all the systems in a complex. This is critical for Parallel Sysplex operation.

2.8.1 Sysplex Timer

A Sysplex Timer® provides the synchronization for the time-of-day (TOD) clocks of multiple servers, and thereby allows events started by different servers to be properly sequenced in time. When multiple servers update the same database, all updates are required to be time stamped in proper sequence.

The zSeries servers can attach to either an IBM 9037 Model 1 or Model 2 Sysplex Timer Unit.⁶ More information on the IBM 9037 can be found in the IBM Redbook *S/390® Timer Management and IBM 9037 Sysplex Timer*, SG24-2070.

ETR attachment

A server's External Time Reference (ETR) feature provides the interface to the IBM 9037 Sysplex Timer. The z9-109 server ETR feature is optional. When it is installed, it provides ETR connections to one or two IBM 9037 Sysplex Timer units.

The two z9-109 server ETR features (when installed) are located in the processor cage of the z9-109 server. Each feature has a single port supporting an MT-RJ fiber optic connector to provide the capability to attach to a Sysplex Timer unit. The 9037 Sysplex Timer unit has an optical transceiver that supports an ESCON Duplex connector.

Note: The z9-109 server ETR feature does not support a multimode fiber optic cable terminated with an ESCON Duplex connector. However, 62.5 micron multimode ESCON Duplex jumper cables can be reused to connect to the ETR feature. This is done by installing an MT-RJ/ESCON conversion kit between the ETR feature MT-RJ port and the ESCON Duplex jumper cable.

⁶ Note that the IBM 9037 model 1 was withdrawn from marketing and service support was discontinued at the end of 2003. IBM intends to withdraw the IBM 9037 model 2 from marketing in 2006. When the withdrawal from marketing is announced, no order for a Sysplex Timer or its features will be accepted after the effective date.

2.8.2 Server Time Protocol (STP)

STP is an important new time synchronization function for the z9-109 server (also the z990 and z890 servers), and is being previewed with the z9-109 server. This function coordinates the time-of-day clocks in a multiple z9-109, z990, or z890 environment, *whether or not they are part of a Parallel Sysplex*. In configurations consisting only of z9-109, z990, and z890 machines the new function will eliminate the requirement for IBM 9037 Sysplex Timers to coordinate the time. If non-STP capable servers are included in the timing network, then the IBM 9037 Sysplex Timers are still required

This time synchronization function has the following two characteristics:

- ▶ *Synchronization* of the time-of-day clocks across multiple systems, within the constraints involved for consistent timestamps in various applications.
- ▶ Having the capability to *set the time to an international time standard*, such as Universal Coordinated Time (UTC), and periodically adjusting the time so that it stays within a certain tolerance of UTC after it is set.

The z9-109 server (and also z990 and z890 servers) will be STP-capable and be able to send timing information through ISC-3 links configured in *Peer mode*, ICB-4, or ICB-3 links. These may be the same links used for Coupling Facility operation; the link will be shared automatically. If such links are not present (when Parallel Sysplex is not involved) then the links must be installed for use by STP. The ISC-3 link can operate at distances of up to 100 km (62 miles).

If there is a requirement to set the server time to UTC, STP allows you to dial out to an external time service to obtain an accurate time reference and set the time. You can then select to periodically dial out and adjust the time in order to maintain time accuracy.

STP will also allow you to set time zone and Daylight Savings Time *offsets*. The ability to automatically update Daylight Savings Time is also provided.

2.9 Hardware Storage Area

The Hardware Storage Area (HSA) in the z9-109 server may reside in any book (or split across books) and may be automatically shifted if the server is restarted in a reduced configuration.

HSA memory is used for several internal functions, but the bulk is used for I/O control blocks. Each subchannel in each LPAR requires 256 bytes in HSA. The HSA *could* be substantially larger on the z9-109 server than on earlier systems. This is directly related to the Multiple Subchannel Set function that provides approximately twice as many I/O devices as provided by earlier systems. An HSA estimation tool is available through the IBM Resource link at:

<https://www-1.ibm.com/servers/resourceLink/hom03010.nsf/pages/tools?OpenDocument&login>

2.10 System control

Figure 2-8 provides a conceptual overview of the system control design. While some details have changed, the general structure for system control is similar to the z990.

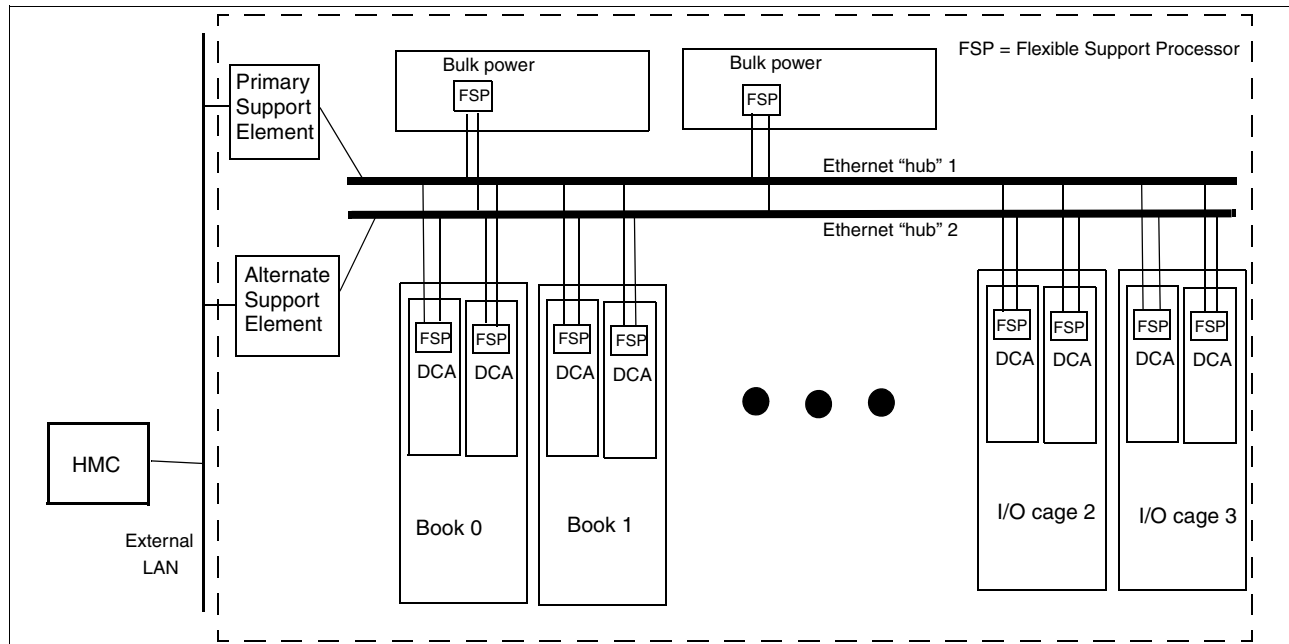


Figure 2-8 System control

Various system elements contain *Flexible Support Processors (FSPs)*.⁷ An FSP is based on the IBM Power PC® microprocessor. It connects to an internal Ethernet LAN (to communicate with the Support Elements) and provides a SubSystem Interface (SSI) for controlling components.⁸

A typical FSP operation is to control a power supply (shown as DCAs in the figure). A Support Element might send a command to the FSP to bring up the power supply. The FSP (using the SSI connections) would cycle the various components of the power supply, monitor the success of each step, monitor the resulting voltages, and report this status to the Support Element.

A z9-109 server has more FSPs than indicated in the figure and the interconnections are more complex, but Figure 2-8 illustrates the general concepts involved. Most system elements are duplexed (for redundancy) and each element has an FSP. There are two internal Ethernet LANs and two Support Elements, again for redundancy. There is a crossover capability between the LANs, so that both Support Elements can operate on both LANs.

The Support Elements, in turn, are connected to another (external) LAN (Ethernet only) and the Hardware Management Consoles (HMCs) are connected to this external LAN. There can be one or more HMCs. In a production environment, the server complex is normally managed from the HMCs. If necessary, the system can be managed from either Support Element. Several or all HMCs can be disconnected without affecting system operation.

2.11 HMC and SE

The Hardware Management Console and Support Element functions of the z9-109 server have a completely new implementation. OS/2 is no longer used as a base operating system and both are *closed* systems. This means that other applications or user programs may not

⁷ These were informally known as *cage controllers* in earlier systems.

⁸ More detailed descriptions may refer to SSI-M and SSI-S notations, corresponding to master and slave functions.

be added to these systems. This prohibition includes ESCON Director control programs and Parallel Sysplex Timer controls; separate PCs are now needed for these functions. The purpose and general use of the HMCs and SEs remain unchanged, but the appearance of various panels has changed.

The HMC user interface is now implemented through a standard Web browser. Remote users may connect to the HMC using a browser, and have complete HMC functionality.

2.12 Concurrent maintenance and upgrades

The z9-109 server has advanced the ability to permit concurrent maintenance and upgrades. In particular, it provides the following:

- ▶ Server books may be added concurrently with server operation.
- ▶ In some cases, a server book may be removed concurrently with server operation. This is a sophisticated process that requires appropriate planning. In support of this, the z9-109 server can dynamically remap memory while a server is running, allowing the memory on a specified book to be isolated prior to removing the book.
- ▶ In specified cases, new levels of driver code may be installed concurrently with server operation.

2.13 Internal Battery Feature

The Internal Battery Feature is optional on the z9-109 server. In the event of an interruption to the input power, the internal battery will provide sustained server operation for a short period of time. The duration of the battery support is highly dependent upon the server model and I/O cages installed.

Many power interruptions are due to the temporary loss of a single phase on the three-phase input line. The z9-109 server can tolerate loss of a single phase even with no battery backup. This capability, coupled with the Internal Battery Feature and full operation from either of the dual independent line cords, gives a very high degree of resilience to transient power dropouts. When coupled with a UPS capability, the internal battery ensures that there is no loss of power during startup of the emergency supply.

Batteries are installed in pairs; a single ordered feature code (FC 3210) provides two batteries. The pair consists of a front and rear battery in one of the three IBF slots. The Z frame has two IBF slots (at the top of the frame) and the A frame has a single IBF slot (at the top).

2.14 Power and cooling

The z9-109 server footprint is the same as that of a z990 and very similar to a z900. The power service specifications are the same, but the power consumed is slightly greater.

Table 2-5 Power and cooling

Model	One I/O Cage		Two I/O Cages		Three I/O Cages	
	Power	Cooling	Power	Cooling	Power	Cooling
S08	6.3 KW	21.5	9.2 KW	31.4	12.1 KW	41.3

	One I/O Cage		Two I/O Cages		Three I/O Cages	
Model	Power	Cooling	Power	Cooling	Power	Cooling
S18	8.8 KW	30.0	11.8 KW	40.2	14.7 KW	50.1
S28	10.9 KW	37.2	13.9 KW	47.4	16.9 KW	57.6
S38	12.8 KW	32.6	15.7 KW	53.5	18.3 KW	62.4
S54	12.8 KW	32.6	15.7 KW	53.5	18.3 KW	62.4

The cooling figures are thousands of BTUs per hour. These figures assume maximum I/O cage configurations and should be considered approximate.

A z9-109 server uses three-phase 200-480 VAC power. A small S08 server uses approximately 6.3 KW (if all I/O feature slots are filled), while a very large S38 server uses up to 18.3 KW. Two power feed cables are connected to two internal bulk power units that provide power for the server. The server can continue operation after a failure of one of these units. They are designed such that the server will continue to operate after a failure of one phase of the incoming power supply.

Internal functions adjust the power factor to almost 1.0, with little harmonic content. (This complies with increasing demands, sometimes in the form of binding requirements, for *building-friendly* power supplies.)

Incoming power is converted to high voltage DC, which is distributed to the subsidiary power units that convert it to the various voltages needed by the controllers, cages, books, and I/O features. Very large currents are needed at low voltages and it is not practical to directly distribute the low voltage supplies throughout the server.

As an overview statement, everything in the power system is duplexed. Each book (up to four) exists on its own power boundary, meaning that it is possible to remove power on a book without affecting other books.⁹ Two additional power supply units are installed in the processor cage as part of an upgrade to the next higher model.

Units known as DCAs convert the distributed high voltage to various low voltages. Slightly different DCAs are used in the processor cage and the I/O cages. Up to eight DCAs are used for the processor cage (two for each book), and each I/O cage contains two DCAs. Each DCA has a Flexible Support Processor that is part of the system control function discussed in "System control" on page 22.

2.14.1 Cooling

The z9-109 server uses refrigeration cooling for the MCMs in the books. Everything else is air-cooled.

The refrigeration package is a Modular Refrigeration Unit (MRU). A single MRU has two cooling loops and cools the MCMs in one or two books. A z9-109 server model S28 or larger will have two MRUs. The MRUs are not redundant. If an MRU fails such that MCM temperatures rise above specified levels, the system cycle time is automatically slowed down. (The cycle time is the same for all books, and a slowdown affects all books.) MCMs produce less heat when run slower, and an integrated backup air-cooling mechanism (with fans) is sufficient at the slower speed.

⁹ There are many implications to removing a book. This discussion is only about the power aspects.

System slowdown (after an MRU failure) is in steps, depending on the temperatures in the books.

The two cooling loops in an MRU are controlled independently. For example, a book connected to one cooling loop can be removed or added without affecting the operation of another book being cooled by the other loop.¹⁰

The other components of the system (I/O cages, power supplies, memory in the books, and so forth) are air-cooled. A sophisticated arrangement of fans (Air Movement Devices, or AMDs) and sensors provides the operation. Fan speeds are variable and are controlled to maintain the required cooling.

The total heat produced by the z9-109 server ranges from approximately 21,500 BTUs to 62,400 BTUs. This heat load must be managed by the raised-floor air system.

¹⁰ This only applies to the cooling function, removing or adding a book involves many other considerations.



Key functional elements

This chapter discusses various topics of special interest in a z9-109 server. Some topics are completely new, such as the multiple subchannel sets, while others were introduced with the z990 or z890 servers.

The topics discussed in this chapter include:

- ▶ Availability enhancements
- ▶ Book removal
- ▶ Enhanced driver maintenance
- ▶ Redundant I/O Interconnect
- ▶ Logical channel subsystems
- ▶ Multiple subchannel sets
- ▶ OSA for NCP
- ▶ GARP VLAN Registration Protocol (GVRP)
- ▶ The MIDAW facility
- ▶ N_Port ID virtualization
- ▶ New and changed instructions
- ▶ Branch detection
- ▶ Performance factors
- ▶ Program-directed re-IPL
- ▶ Open exchanges for FCP and FICON
- ▶ CHPID mapping
- ▶ LPARs
- ▶ Hardware Management Console and Support Elements

3.1 Availability enhancements

The z9-109 servers include several availability enhancements. Several of these are discussed separately throughout this document. For completeness, we list the primary availability enhancements here as a group:

- ▶ *Redundant I/O interconnect* groups I/O cage domains into pairs and allows the STI link to each domain to service the other half of the pair if the STI link to that domain fails or is not operational for some reason. When possible, the two STI links to the paired domains should be from different books since this provides the best availability configuration. The STI elements (also known as MBA connections) may be removed and replaced concurrently with server operation. Each element contains two STI connections.
- ▶ *Dynamic oscillator switchover* allows the backup system oscillator to transparently take over system clocking if the primary oscillator fails. Earlier servers had a backup oscillator but a power-on reset (POR) was needed to use it.
- ▶ The *flexible memory option* uses additional memory (ordered by the customer) to be spread throughout the server books such that, if a book is unavailable, the server will still have its intended amount of memory available. That is, the additional memory is held in reserve and is automatically invoked when needed to maintain the intended operational size of memory.
- ▶ *Enhanced book availability* allows a book to be removed from a server concurrently with server operation (in a multi-book system, of course). This requires detailed advance planning and operational management. A book may be added in a corresponding manner. This remove/add function can be used, for example, to add memory to an existing book concurrently with server operation. Note that this function is for the *planned* removal of a book. An abrupt book failure will require a system restart during which the server will *fence* the failing book and that book can then be removed (concurrently with server operation using the remaining books).
- ▶ *Enhanced driver maintenance* allows selected driver upgrades to be installed concurrently with normal operation. This may be possible when moving from selected *sync points* of driver n-1 to a corresponding sync point in driver n. This function depends on the driver sync points provided by IBM and may not be available for some upgrades.
- ▶ The *memory sparing* function used with the memory design of the z9-109 server provides more spare chips on the memory cards than were used in earlier designs. The design makes it very unlikely that DRAM failures will affect server operation.
- ▶ Improved displays for *FICON link status* provides additional information for debugging FICON connection issues.
- ▶ *FICON link recovery thresholds* (a software function) reduce the potentially severe software overhead created by multiple link failures.

Other availability designs have been carried forward from zSeries machines. These include:

- ▶ Every ESCON adapter card has a spare channel port and this may be configured and used without disrupting server operation.
- ▶ All I/O cards (as well as the new Crypto Express2 card) may be removed or installed concurrently with server operation, with appropriate planning.

3.2 Book removal

Removing a book, concurrently with server operation, requires careful planning and assumes that sufficient resources (processors, memory, STI links) for meaningful operation will exist

without the targeted book. A very important function has been included with the z9-109 server to assist removal of a book. This is the ability to *remap* the real memory used by LPARs in order to move LPAR memory from the selected book to other books.¹¹ This remapping can be done concurrently with server operation. This assumes that sufficient memory exists in the remaining books, of course. The flexible memory option is intended for this purpose. Alternatively, it may be possible to delete less important LPARs to free sufficient real memory for the remapping.

As part of the removal process the server will activate whatever unused PUs are available in the remaining books to attempt to provide the configured number of PUs for server operation.¹² If redundant I/O interconnect was properly planned (with STI pairs connected to two different books) it should be possible to remove a book without losing any I/O connectivity. Any ICBs (ICB4 or ICB3) must be configured offline from the affected book.

3.2.1 Book replacement

Book replacement, done concurrently with server operation, has been possible on z990 servers. The replaced book might have additional memory added to it. The z9-109 server has improved the earlier book replacement processes by better automating the reconfiguration of PUs (to use the new book) and the remapping of physical memory for best LPAR positioning.

3.3 Enhanced driver maintenance

Some server driver¹³ transitions can be installed concurrently with server operation. These concurrent transitions occur only at IBM-defined *sync points*, as shown in Figure 3-1.

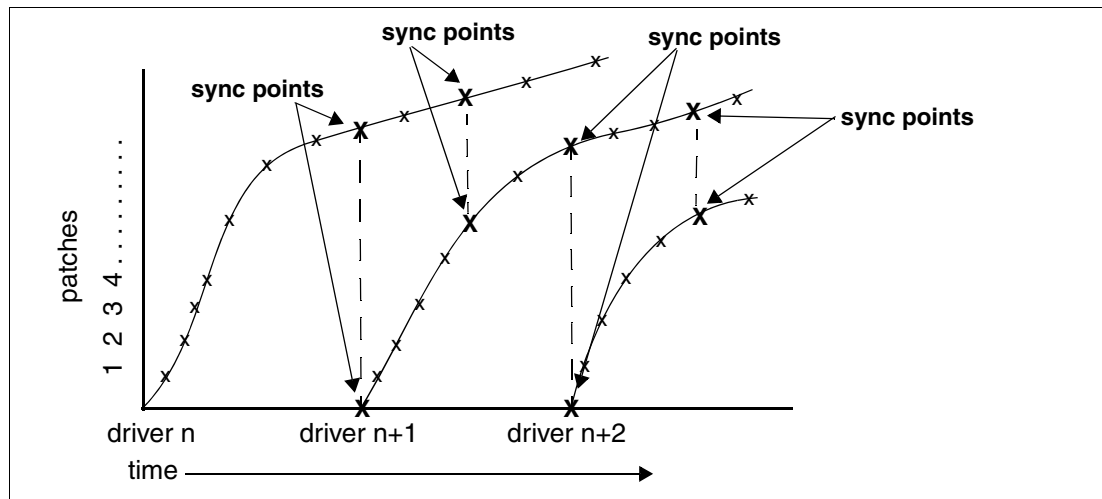


Figure 3-1 Driver sync points

The sync points are defined only at selected patch levels of various drivers. Concurrent transitions are available only for single-level jumps; that is, from driver n to n+1. A concurrent

¹¹ This is a substantial technical advance that makes taking a large “bank” of memory offline a practical possibility. The real and absolute address of the memory is unchanged. A lower level of memory mapping (using addressing techniques not visible to software) moves 64 MB blocks of memory between books. I/O to memory within the block is temporarily paused while the memory is moved and remapped. This is all completely transparent to software.

¹² This may include PUs (or memory) that are physically installed but have not been purchased.

¹³ A “driver” is the IBM Licensed Internal Code that provides many server functions that are generally regarded as “hardware” functions.

jump from driver n to $n+2$, for example, is not available. Also, a concurrent move to a previous driver level is not possible.

IBM generally intends to define a limited number of sync points (“crossover points”) for each driver. However, this does not preclude the potential occurrence of disruptive driver upgrades; these will be documented as far in advance as possible.

Certain manual actions may be needed after a concurrent driver upgrade. These include the following:

- ▶ Non-QDIO OSA-type CHPIDs must be varied off and varied on again to activate any new code. In particular, this applies to CHPID types OSE and OSC.
- ▶ If new basic FICON adapter LIC is present in the new driver then FICON and FCP CHPIDs must be reset to activate the new code.

3.4 Redundant I/O Interconnect

STI ports (used for connections between a book and an I/O cage domain connections) are installed in pairs and pairs may be installed or removed concurrently with server operation. STI port removal assumes, of course, that there is sufficient I/O connectivity or configuration flexibility to withstand the removal of the STI connections.

Furthermore, I/O cage domains are now paired (with the last domain acting as its own pair). The two STI connections to the two domains (in a pair) provide automatic backup for each other. If one of the STI connections fails or is removed, the STI connection for the other member of the pair will handle all I/O adapters in both domains. In this case the potential exists that the data rate of all the adapters may overrun the bandwidth of the single STI connection. In this case, some I/O operations may be slightly delayed.

The redundant I/O interconnect function does not apply to ICB4 or ICB3 connections.

3.5 Logical channel subsystems

One of the most striking architectural changes associated with the z9-109 server (and the z990 and z890 servers) is the set of extensions to the channel subsystem. Perhaps the best way to start a description is to explain *why* these extensions were introduced. The purpose is simple; it is to permit a server to have more than 256 channels. The maximum number of channels has been a concern for some time, although the move to FICON channels and devices has mitigated the limitation.

The difficulty is to extend this number while maintaining compatibility with existing software. The 256 maximum CHPID number is reflected in various control blocks in operating systems, software performance measurement tools, and even some application code. Simply changing the control blocks is not a viable option since this would break too much existing code. The solution provided introduced with the z990 is in the form of logical channel subsystems (LCSSs). These are implemented in a manner that has little or no impact on existing code.

It is important to note that there is no unique hardware (or feature codes) associated with LCSSs. These are not priced features. These architected functions are implemented in the LIC of the server and are, potentially, exploited by operating systems running in the server.

Existing code works with single-byte CHPID numbers, producing the limitation of 256 CHPIDs. The logical channel subsystems provides *multiple sets* of channel definitions, each with a maximum of 256 channels. Existing operating systems would be associated with one

logical channel subsystem (LCSS) and work with a maximum of 256 CHPIDs. Different LPARs can be associated with different LCSS definitions. Thus a single operating system instance (using existing code) still has a maximum of 256 CHPIDs, but the server as a whole can have more than 256 CHPIDs. It is necessary to have multiple operating images (in multiple LPARs) to exploit more than 256 channels, but this is a common mode of operation.

In a sense, LCSSs virtualize CHPID numbers. A CHPID no longer directly corresponds to a hardware channel, and CHPID numbers may be arbitrarily assigned. A hardware channel is now identified by a PCHID, or *physical* channel identifier. A PCHID number is defined for each *potential* channel interface. An I/O adapter card has up to 16 channel interfaces¹⁴ and 16 PCHID numbers are reserved for each I/O adapter slot in each I/O cage. Not all I/O adapters provide 16 channels, of course, but 16 PCHID numbers are allocated to each I/O slot. The PCHID numbers allocated to each I/O adapter and port on that adapter are fixed¹⁵ and cannot be changed by the user.

The z990 and the z9-109 server can use up to four logical channel subsystems, although the architecture permits more than four LCSSs. More complete exploitation of this architecture is expected to be delivered over time, in evolutionary steps, by the operating systems and various software subsystems.

A given LPAR is associated with a single LCSS, and a single LCSS has a maximum of 256 CHPIDs. Multiple LPARs may be associated with a given LCSS. Using two LCSSs means that up to 512 channels can be used, and so forth. This number is reduced by *spanned* channels and these are discussed in “Spanned channels” on page 32.

Figure 3-2 illustrates the relationship between LPARs, LCSSs, CHPIDs, and PCHIDs. This is an idealized illustration and ignores such complexities as spanned channels, dynamic I/O changes, and so forth.

¹⁴ No existing I/O adapter provides 16 usable interfaces (channels). Note that an ESCON adapter has 16 interfaces but only 15 may be used; the 16th is a spare.

¹⁵ There is a minor exception for ESCON adapters when a spare port replaces a failing port.

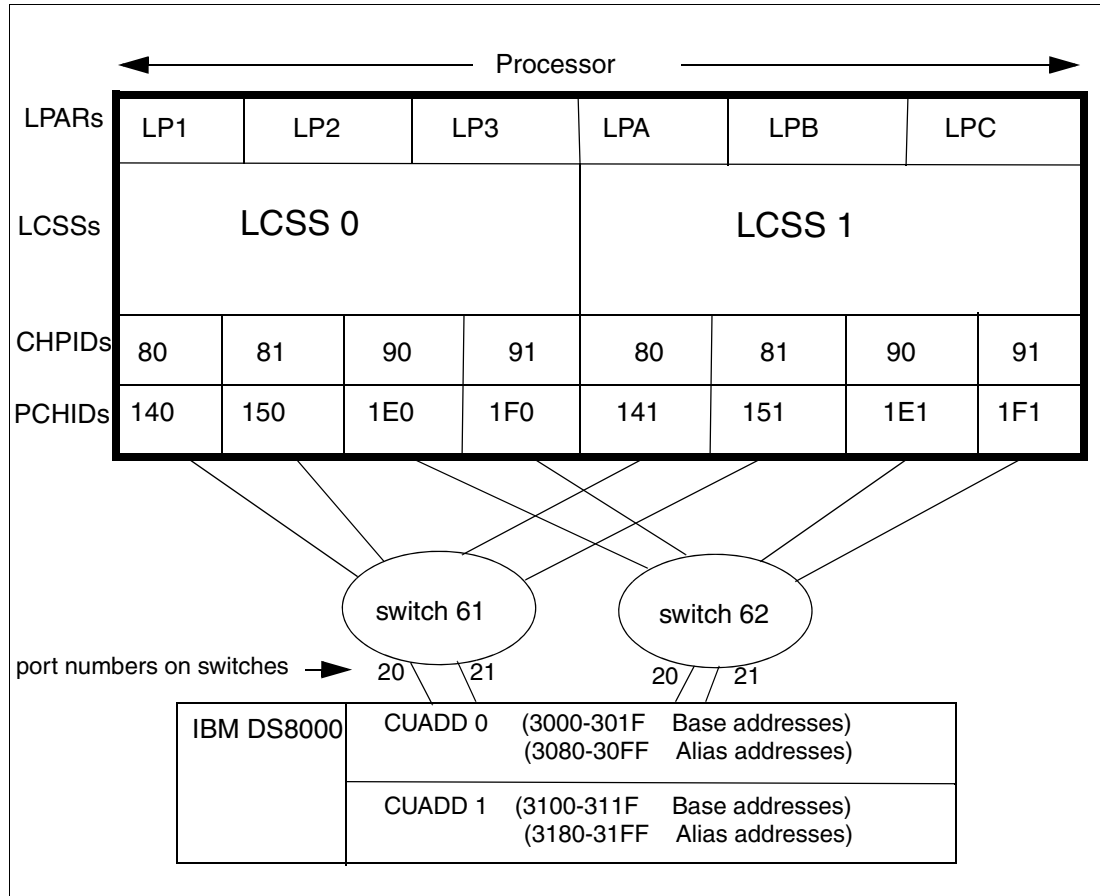


Figure 3-2 Overview of LPARs, LCSSs, CHPIDs, and PCHIDs

This figure illustrates a number of important points, such as:

- ▶ Two CSSs are shown.
- ▶ Every LPAR is associated with a specific CSS. (Also note that LPARs have unique names across the complete server. LPAR naming and numbering has become a little more complex.)
- ▶ Multiple LPARs (up to 15) may be associated with a CSS.
- ▶ A CHPID is associated with a specific CSS. CHPID numbers are unique within that CSS, but may be reused in other CSSs. (For example, there is a CHPID 80 in both CSSs.) A CHPID number is arbitrarily selected. For example, we could change CHPID 80 (in either or both CSSs in the illustration) to C3 simply by changing a value in the IOCDs.
- ▶ A CHPID is associated with a PCHID, and PCHID numbers are unique across the server.
- ▶ Different channels on a single I/O adapter can be used by different LPARs. In the illustration, PCHID 0140 is the first channel on the adapter in I/O cage 1, slot 6. PCHID 0141 is the second channel on the same adapter. See "Physical I/O connections" on page 18 for a discussion of PCHID values and I/O cage slots.

3.5.1 Spanned channels

Spanning channels is the ability to configure a channel to more than one logical channel subsystem. The channel is transparently shared by any or all of the configured logical partitions regardless of the LCSS to which the partition is configured.

Spanned channel support is an extension of Multiple Image Facility (MIF). MIF allows sharing of channels across LPARs and MIF Spanned channel allows sharing across LCSSs. When

LCSSs were introduced, transparent sharing of internal channels was introduced—sharing of HiperSockets and Internal Coupling Channels (ICs) between CSSs. Support on the z9-109 server (and the z990 and z890 servers) now includes sharing of internal channels (HiperSockets and ICs), as well as sharing of external channels—FICON Express (FC and FCP), ICB-3, ICB-4, ISC-3 (CFP and CFS), and OSA-Express features; they can all now be configured as MIF spanning channels, allowing sharing of channel resources across Logical Partitions.

Spanned channels can be shared among LPs across LCSSs. A spanned channel occupies the same CHPID number in all LCSSs in which it is used. For example, if a HiperSocket is to be used across two LPARs belonging to two different LCSSs, then the CHPID must be the same in both LCSSs. A HiperSocket that connects LPARs in different LCSSs *must* be spanned. Table 3-1 indicates possible spanning and sharing combinations.

Table 3-1 Spanned and shared channels

Channel type		CHPID definition	MIF shared channel	MIF spanned channel
ESCON	External	CNC, CTC	Yes	No
		CVC, CBY	No	No
FICON	External	FC, FCP	Yes	Yes
		FCV	Yes	No
OSA	External	OSC, OSD, OSE, OSN	Yes	Yes
ICB-4	External	CBP	Yes	Yes
ICB-3	External	CBP	Yes	Yes
ISC-3	External	CFP	Yes	Yes
IC	Internal	ICP	Yes	Yes
HiperSockets	Internal	IQD	Yes	Yes

3.5.2 Channel definitions in the IOCP statement

The following channel types (as defined in an IOCDs) are used with these servers:

- ▶ FICON channel types
 - FC - Native FICON channel (both for native FICON devices and FICON CTCs)
 - FCV - FICON bridge channel (FICON Express LX only)
 - FCP - Fibre Channel Protocol (SCSI)
- ▶ ESCON channel types
 - CNC - Native ESCON channel
 - CTC - ESCON CTC channel
 - CVC - ES conversion channel, which connects to a converter in block multiplexer mode
 - CBY - ES conversion channel, which connects to a converter in byte multiplexer mode
- ▶ CF link channel types
 - CBP - Integrated Coupling Bus (ICB-3) channel, for both OS and CF partitions, to connect z900 or z800 servers
 - CBP - An ICB-4 connection to another z990 or z9-109
 - CFP - InterSystem Coupling (ISC-3) peer mode channel, for both OS partitions and a CF partition
 - CFS - ISC-3 compatibility mode sender channel, for OS partitions¹⁶

- CFR - ISC-3 compatibility mode receiver channel, each must be used for only 1 CF partition
- ICP - Peer mode Internal Coupling (IC-3) channel, for both OS and CF partitions, to connect among LPARs within a z990 server internally
- ▶ OSA-Express channel types
 - OSD - (QDIO operation)
 - OSE - (non-QDIO operation)
 - OSC - (3174 emulation for consoles and VTAM® local 3270s)
 - OSN - (ESCON emulation for emulated 3745-type functions via CCL NCP)
- ▶ HiperSockets channel type
 - IQD - HiperSockets channel, QDIO mode only

Each of these channel types requires that a CHPID be defined, even if it is an internal channel and no physical hardware (channel card) exists. Each channel, whether a “real” channel device or a virtual device (such as a HiperSocket) must be assigned a unique CHPID within the CSS. There are no default CHPID numbers and you can arbitrarily assign whatever number you like (within the X'00' to X'FF' range, of course).

Most of these channel types can be shared and used concurrently among multiple LPARs within the *same* CSS. This capability is known as the Multiple Image Facility (MIF). Exceptions are for ES conversion channels (CVC and CBY) and CF receiver channels (CFR). These channel types cannot be shared concurrently, but can be defined as reconfigurable channels by specifying the REC parameter on the channel definition. The channel can be reassigned to another LPAR after the former owning LPAR configures the channel offline.

CHPID types BY, BL, CBR/CBS, CFR/CFS, or OSA are not allowed.

3.6 Multiple subchannel sets

The *multiple subchannel sets* (MSS) functionality is completely new with the z9-109 server and should not be confused with multiple channel subsystems. In most cases a subchannel represents an addressable device. A disk control unit with 30 drives uses 30 subchannels (for base addresses), and so forth. An addressable device is associated with a device number and the device number is commonly (but incorrectly) known as the device address.

Subchannel numbers (including their implied path information to a device) are limited to four hexadecimal digits by hardware and software architectures. Four hexadecimal digits provides 64K addresses, known as a *set*. IBM reserved 1024 subchannels, leaving 63K subchannels for general use.¹⁷

Again, addresses, device numbers, and subchannels are often used as synonyms although this is not technically correct. We may hear that there is “a maximum of 63K addresses” or “a maximum of 63K device numbers.”

The advent of Parallel Access to Volumes (PAV) has made this limitation (63K subchannels) a problem for larger installations. A single disk drive (with PAV) often consumes at least four subchannels.¹⁸ It was difficult to remove this constraint because the use of four hexadecimal

¹⁶ When you use z/OS system-managed CF duplexing, one CF partition can additionally share the CF sender channel.

¹⁷ The number of reserved subchannels is changed from 1024 to 256, starting with the z9-109 server. We abbreviate this to 63K in this discussion to easily differentiate it from the 64K subchannels available in subchannel set 1.

¹⁸ Four appears to be a popular number for PAV. It represents the base address and three alias addresses.

digits for subchannels (and device numbers corresponding to subchannels) is architected in a number of places. Simply expanding the field would break too many programs.

The solution is to allow *sets* of subchannels (“addresses”), with a current implementation of two sets. Each set provides 64K addresses. (Subchannel set 0, the first set, still reserves subchannels for IBM use although the number of reserved subchannels is being reduced from 1024 to 256). Subchannel set 1 provides a full range of 64K subchannels.

In principle, subchannels in either set could be used for any device addressing purpose. However, the current implementation (in z/OS) restricts subchannel set 1 to disk *alias* subchannels. Subchannel set 0 may be used for base addresses and for alias addresses.

There is no required correspondence between addresses in the two sets. For example, we might have device number 8000 in subchannel set 0 and device number 8000 in subchannel set 1 and they might refer to completely separate devices. (We know that the device number in subchannel set 1 must be an alias for z/OS, but that is all we can know from the device number.) Likewise, device number 1234 (subchannel set 0) and device number 4321 (subchannel set 1) might be the base and an alias for the same device. There is no *required* correspondence between the device numbers used in the two subchannel sets.

An astute reader may realize that the additional subchannel set, in effect, adds an extra high-order digit (either 0 or 1) to existing device numbers. For example, we might think of an address as 08000 (subchannel set 0) or 18000 (subchannel set 1). This is not done in system code or in messages because of the architectural requirement for four-digit addresses (device numbers, subchannels). However, some messages will contain the subchannel set number and one can mentally use it as a high-order digit for device numbers. There should be few requirements for this since subchannel set 1 is used only for alias addresses and users (JCL, messages, documentation, programs) seldom refer directly to an alias address.

Note that *each* logical channel subsystem (LCSS) can have multiple subchannel sets, as outlined in Figure 3-3.

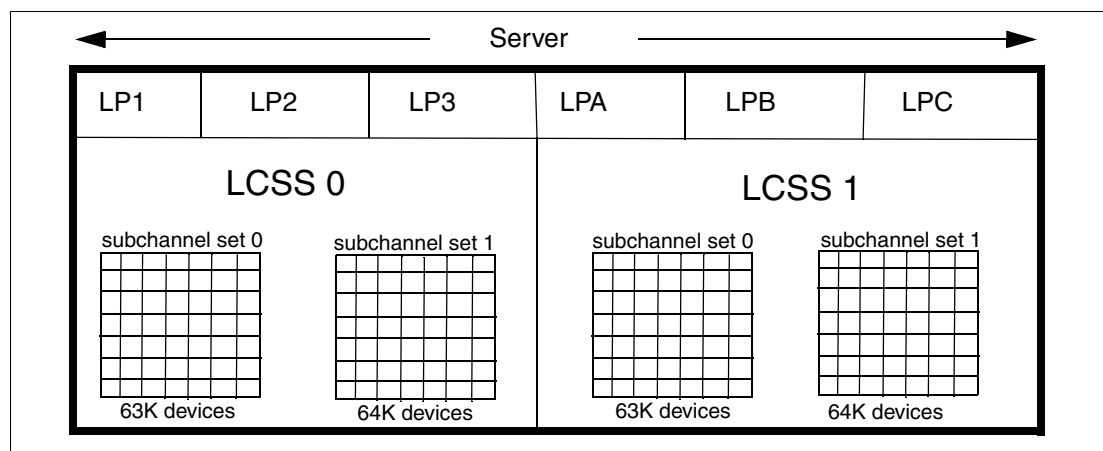


Figure 3-3 Multiple logical channel subsystems (LCSS) and multiple subchannel sets

The appropriate subchannel set number must be included in IOCP definitions (or in the HCD definitions that produce the IOCDS). The subchannel set number defaults to zero and IOCP changes are needed only when using subchannel set 1. Shown in 3.6.1, “IOCP example” on page 36 is an extract of an IOCP listing that uses both multiple logical channel subsystems (LCSSs) and multiple subchannel sets (MSSs).

3.6.1 IOCP example

LCSS, MSS, and PCHID information is included in an IOCDs. An IOCDs is created through the HCD utility program or by use of a standalone IOCP file. In practical use, HCD is normally used. For illustration purposes we will examine an IOCP file and will assume the reader is generally familiar with such files.

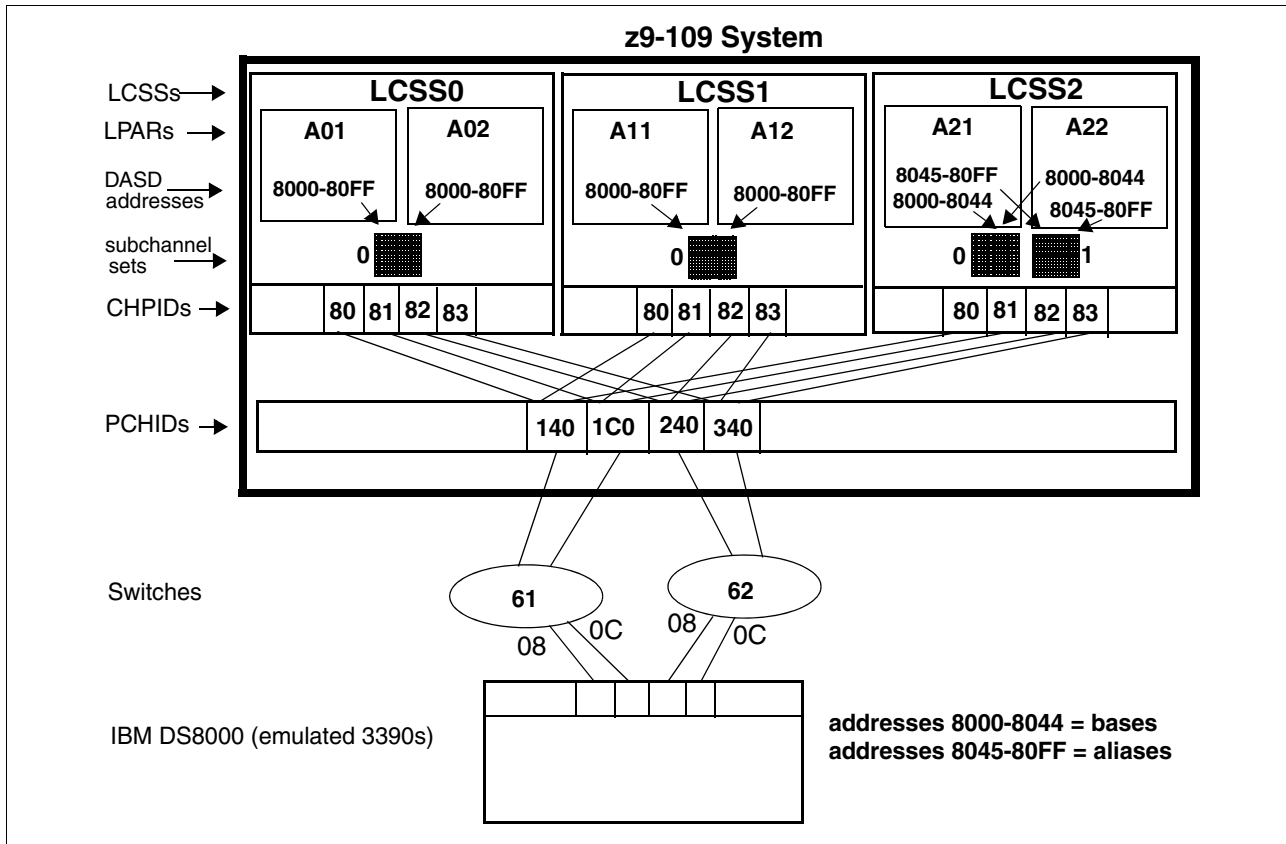


Figure 3-4 IOCP Example with LCSSs, MSSs, and PCHIDs

The following IOCP definition describes the server shown in Figure 3-4. This is not intended to represent a practical server—it has no consoles, for example—but it illustrates the new elements involved.

```

ID      MSG1='IODF01',
        MSG2='HAIMO.IODF01.WORK - 2005-05-23 13:59',
        SYSTEM=(2094,1),
        TOK=('SCZP101',008001136A3A2084135941210105143F00000000,
        00000000,'05-05-23','13:59:41',' ',' ')
RESOURCE PARTITION=((CSS(0),(A01,1),(A02,2)),(CSS(1),(A11,1),
        A12,2)),(CSS(2),(A21,1),(A22,2))),
MAXDEV=((CSS(0),65280,0),(CSS(1),65280,0),(CSS(2),65280,
        65535))
CHPID  PATH=(CSS(0,1,2),80),SHARED,SWITCH=61,PCHID=140,TYPE=FC
CHPID  PATH=(CSS(0,1,2),81),SHARED,SWITCH=61,PCHID=1C0,TYPE=FC
CHPID  PATH=(CSS(0,1,2),82),SHARED,SWITCH=62,PCHID=240,TYPE=FC
CHPID  PATH=(CSS(0,1,2),83),SHARED,SWITCH=62,PCHID=340,TYPE=FC
CNTLUNIT CUNUMBR=8000,
        PATH=((CSS(0),80,81,82,83),(CSS(1),80,81,82,83),(CSS(2),
        80,81,82,83)),UNITADD=((00,256)),
        LINK=((CSS(0),08,0C,08,0C),(CSS(1),08,0C,08,0C),(CSS(2),

```

```

08,0C,08,0C),CUADD=0,UNIT=2105
IODEVICE ADDRESS=(8000,069),CUNUMBR=(8000),STADET=Y,UNIT=3390B
IODEVICE ADDRESS=(8045,187),CUNUMBR=(8000),STADET=Y,          *
SCHSET=((CSS(2),1)),UNIT=3390A

```

Key elements in this IOCP include the following:

- ▶ Three logical channel subsystems (CSS0, CSS1, CSS2) are defined. The LCSSs used in the IOCDs created from this IOCP are stated in the RESOURCE statement. This *defines* the LCSSs. Note that LCSS is abbreviated to CSS in the IOCP.
- ▶ Two LPARs are defined in each channel subsystem (A01, A02, and so forth). These are also defined in the RESOURCE statement.
- ▶ The maximum numbers of subchannels (for the relevant channel subsystems) are defined in the RESOURCE statement. CSS0 and CSS1 have the single, default subchannel sets (with a maximum of 65280 devices that may be defined). CSS2 has two subchannel sets (with 65280 for the first and 65535 for the second as the maximum number of devices).
- ▶ Four spanned channels are used (PCHIDs 140, 1C0, 240, 340). Since these are spanned (across LCSSs), they must have the same CHPIDs in each LCSS. They were assigned 80, 81, 82, and 83 in this example. Each CHPID statement specifies its LCSS number as part of the PATH parameter.
- ▶ Each CHPID statement *must* include a PCHID parameter to associate the CHPID with a physical channel, except for internal CHPID types such as ICP and IQD. (The PCHID parameters can be added to the IOCP definitions by the CHPID Mapping Tool, through HCD definitions, or defined in an IOCP input file, but the PCHID parameters must be present in all CHPID statements in order to create an IOCDs.)
- ▶ PATH and LINK parameters in CNTLUNIT statements must indicate the LCSS number associated with each path and link.
- ▶ The target DASD system has four channel interfaces and 256 unit addresses. It has 69 base devices defined at addresses 8000-8044, and 187 aliases defined at addresses 8045-80FF.
- ▶ CSS0 and CSS1 use subchannels from subchannel set 0 to access both the base and alias addresses. CSS2 uses subchannels from subchannel set 0 base devices and subchannels from subchannel set 1 to access aliases. The second IODEVICE statement defaults to SCHSET 0 for CSS0 and CSS1. It is specified to use SCHSET 1 for CSS2.
- ▶ This example does not clearly illustrate the device address relief provided by multiple subchannel sets. However, it does exist in CSS2 in the example. Addresses 8045 through 80FF are free in subchannel set 0 and could be used for additional devices. (Addresses 8045-80FF are used for alias addresses in CSS0 and CSS1. The alias addresses have been moved to subchannel set 1 in CSS2.)

As can be seen from this example, the basic concepts and definitions for multiple logical channels subsystems are straightforward and mesh with existing IOCP parameters. Equivalent fields have been added to HCD panels for entering LCSS and PCHID numbers.

3.7 OSA for NCP

The IBM Communication Controller for Linux (CCL) on the z9-109 server provides a replacement for IBM 3745 controllers working with Ethernet LANs. The 3745s are typically controlling SNA networks. A new release of CCL will be available later for the z9-109 server and a new function of OSA-Express2 adapters will be used. Figure 3-5 on page 38 illustrates the general concept.

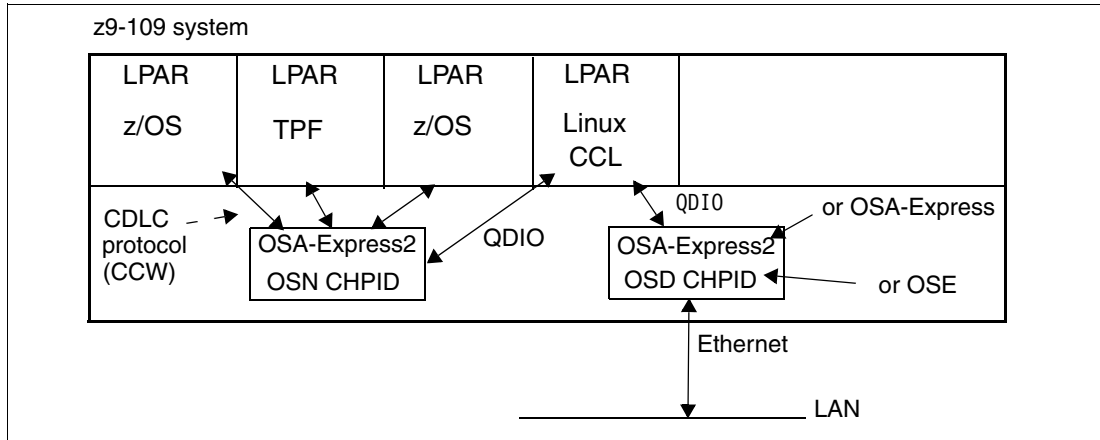


Figure 3-5 General OSA for NCP configuration with the z9-109 server

Two OSA-Express2 ports (on the same or different adapter cards) are involved. One port, operating as a CHPID type OSN appears as an ESCON-connected 3745 interface to the associated operating systems. This OSA-Express2 adapter has, in effect, an internal loop (LPAR-to-LPAR) that sends the data packets to the CCL product running in another LPAR under Linux. The CCL product emulates the appropriate 3745 functions, reformats the data as needed, and sends it on a LAN connected to the second OSA-Express2 port. (The figure and description are presented in terms of outgoing data; incoming data flows in the opposite direction, of course.)

The OSA-Express2 adapter used for the OSN function can be a Gigabit Ethernet LX, Gigabit Ethernet SX, or a 1000BaseT Ethernet adapter; it *cannot* be a 10 Gigabit Ethernet adapter.

3.8 GARP VLAN Registration Protocol (GVRP)

GVRP is defined in the IEEE 802.1P standard for the control of IEEE 802.1Q VLANs. It can be used to help simplify networking administration and management of VLANs.

With GVRP support, an OSA-Express2 port can register or de-register its VLAN IDs with a GVRP-capable switch and dynamically update its table as the VLANs change (see Figure 3-6).

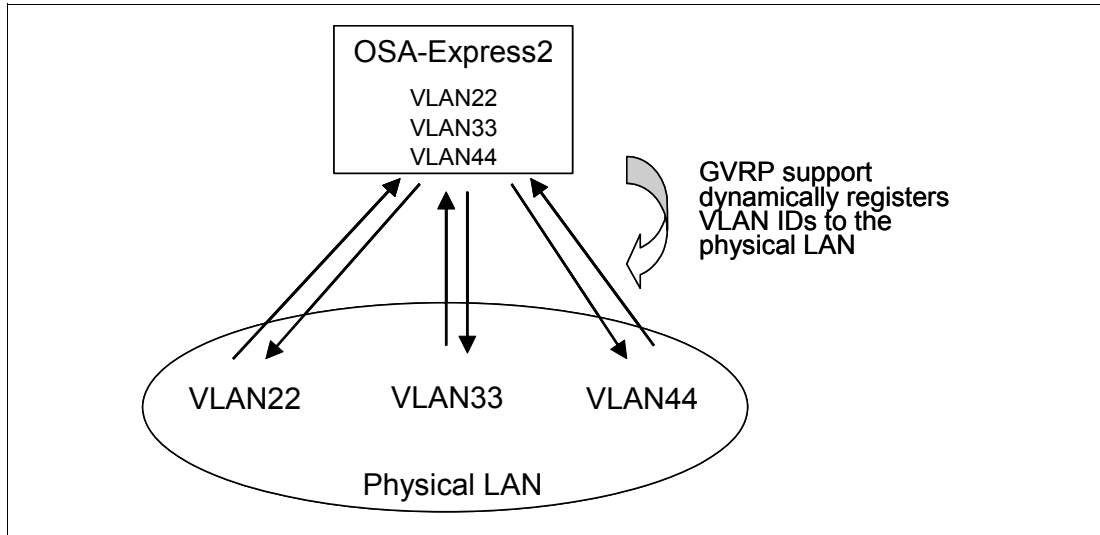


Figure 3-6 GVRP support

Support of GVRP is exclusive to System z9 and applies to all OSA-Express2 features running in QDIO mode (CHPID type OSD).

3.9 The MIDAW facility

The MIDAW function, new with the z9-109 server, provides a more efficient CCW/IDAW structure for certain categories of data chaining I/O operations. The MIDAW name is for Modified IDAW. An IDAW is an Indirect Address Word that is used to specify data addresses for I/O operations in a virtual environment.¹⁹ The existing IDAW design allows the first IDAW in a list to point to any address within a page. Subsequent IDAWs in the same list must point to the first byte in a page; also, all but the first and last IDAW in a list must deal with complete 2K or 4K units of data. This limits the usability of IDAWs to straightforward buffering of a sequential record.

CCWs with *data chaining* may be used to process I/O data blocks that have a more complex internal structure in which portions of the data block are directed into separate buffer areas. (This is sometimes known as scatter-read or scatter-write.) However, data chaining CCWs are inefficient in modern I/O environments for a number of reasons involving switch fabrics, control unit processing and exchanges, numbers of channel frames required, and so forth.

z/OS extended *format* data sets use internal structures (usually not visible to the application program) that require scatter-read (or scatter-write) operation. This means that CCW data chaining is required and this produces less than optimal I/O performance. The use of extended format data sets has grown rapidly, and this has prompted the development of the MIDAW facility.

The modified IDAW (MIDAW) format is shown in Figure 3-7 on page 40. It is 16 bytes long and is aligned on a quadword.

¹⁹ There are exceptions to this statement and we skip a number of details in the following description. We assume knowledgeable readers will merge this brief description with an existing understanding of I/O operations in a virtual memory environment.

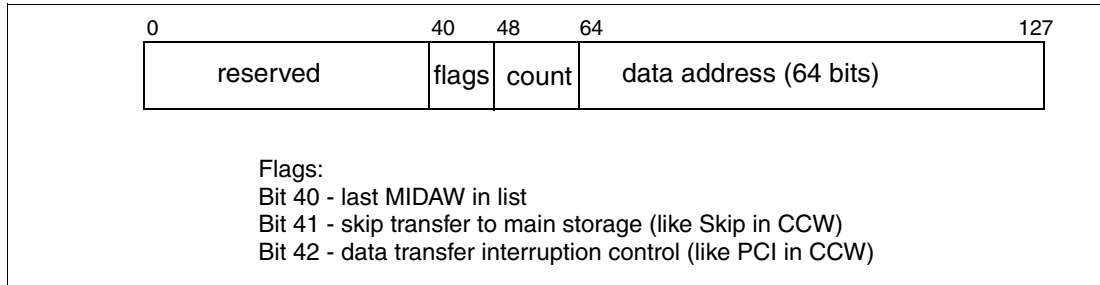


Figure 3-7 MIDAW format

A trivial example of MIDAW usage is shown in Figure 3-8.

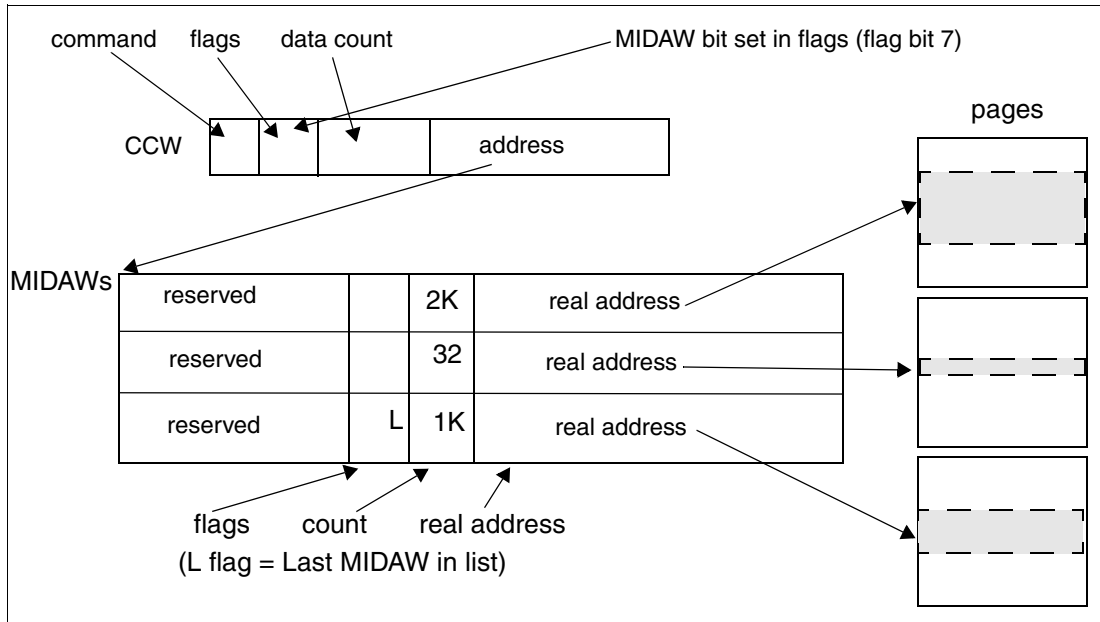


Figure 3-8 MIDAW usage

The use of MIDAWs is indicated by the MIDAW bit in the CCW. If this bit is set, then the *skip flag* may not be set in the CCW. (The skip flag in the MIDAWs may be used instead.) The data count in the CCW should equal the sum of the data counts in the MIDAWs. The CCW operation ends when the CCW count goes to zero or the last MIDAW (with the *last* flag) ends. The combination of the address and count in a MIDAW cannot cross a page boundary; this means the largest possible count is 4K. The maximum data count of all the MIDAWs in a list cannot exceed 64K. (This is because the associated CCW count cannot exceed 64K.)

Notice the scatter-read (or scatter write) effect of the MIDAWs. This makes it possible to efficiently send small control blocks embedded in a disk record to separate buffers than those used for larger data areas within the record. Remember that the MIDAW operations are on a single I/O block, in the manner of data chaining. Do not confuse this operation with CCW *command* chaining.

MIDAWs are used by two IOS drivers:

- ▶ The media manager exploits MIDAWs when appropriate. (The media manager is used to manage low-level I/O functions for extended format data sets and VSAM.)
- ▶ Users of the EXCPVR IOS driver may construct channel programs containing MIDAWs provided they construct an IOBE with the new IOBEMIDA bit set.
- ▶ Users of EXCP *may not* construct channel programs containing MIDAWs.

Note that there is no change to ESCON, FICON, or control unit implementations. However, there is a greater chance that more *information units* (up to 16) may be sent to FICON control units in a single burst.

3.10 N_Port ID virtualization

FICON Express ports may be used for FCP operation instead of FICON operation. This is done by assigning the CHPID type FCP for the channel. FCP operation is used primarily by Linux, but is supported by z/VM and z/VSE™ for some functions.²⁰ This has been available with zSeries servers. This has involved a limitation inherent in the basic FCP design. The Storage Area Network (SAN) regards the FCP channel as a single identity. In SAN terminology, it has a single World Wide Port Number (WWPN).

If the FCP channel is used by multiple independent operating systems (under z/VM or in multiple LPARS), the SAN is unaware of this and cannot distinguish among the various independent users. This has obvious implications for security and file system integrity controls. The problem can be partly avoided by requiring separate FCP channels for each LPAR, but this does not help with z/VM.

A major improvement is provided by N_Port ID virtualization. This allows a single physical FCP channel to be assigned multiple WWPNs and appear as multiple channels to the external SAN environment. In general, a separate WWPN would be provided to each LPAR (not running z/VM) and to each guest in a z/VM LPAR. This permits basic security controls, such as LUN masking and switch-level zoning, to be used by each independent operating system working through the FCP channel. This functionality may be used to permit shared access to the same SCSI LUN in the SAN, although additional considerations are involved in sharing LUNs.

The z9-109 server N_Port ID virtualization is based on a recent extension to the Fibre Channel standards that allows a host bus adapter to perform multiple logins to the SAN fabric via a single physical port. The entry point to the SAN (that is, the switch to which the FCP channel is directly connected) must support *Multiple N-port ID logins*. No changes are needed in any “downstream” switches, bridges, control units, or devices. The switch (and not the FCP channel itself) provides the multiple WWPNs used for virtualization. There are limits on the number of WWPNs that can be provided for a given FCP.

In principle, N_Port ID virtualization can be transparent to software. In practice, new z/VM commands are being provided to query the WWPN names. In addition, software changes for the SCSI driver for Linux will exploit the ability of a switch to provide multiple WWPNs.

²⁰ It is not supported by z/OS, which does not use FCP channels.

3.11 New and changed instructions

The z9-109 server includes a number of new and changed instructions. These are listed in Appendix B, “Instruction set changes” on page 81. These instructions may be grouped into several categories:

- ▶ Hexadecimal floating point instructions for various unnormalized multiply instructions and multiply and add instructions.
- ▶ Several *immediate* instructions, including various add, compare, exclusive OR, OR, subtract, load, and *insert* formats. Use of these instructions, where appropriate, improve performance and can reduce the need for a data base register.
- ▶ Various load instructions, especial those useful for handling unsigned half words (such as used for unicode).
- ▶ The cryptographic instructions (KM, KMC, KIMD, KLMD) have been extended with AES, SHA-256, and random number functions.
- ▶ The *Store facility list extended* instruction can reflect use of more than 32 facilities.
- ▶ The *Extended Translate Facility-2* instructions (TROO, TROT, TRTO, TRTT) have been improved, allowing them to bypass test-character processing when appropriate and relaxing operand-alignment requirements.
- ▶ The *Extended Translate Facility-3* instructions (CU24, CU21, CU12, CU14) have been enhanced to conform with the current Unicode 4.0 standard.

3.12 Branch detection

Debugging a *wild branch* has always been a difficult task. The z9-109 has a new register that contains the address of the last successful branch instruction. If the branch results in an ABEND (or SLIP trap) then the program issuing the branch can be immediately identified.

Starting with z/OS 1.7, this *branch from* information is included in SDWA and RTM2WA dump areas.

3.13 Performance factors

There are several factors that are responsible for the performance characteristics of the z9-109 server processor units. The most obvious factor is that the PU has been designed and optimized to meet the demands of new workloads, while at the same time maintaining excellent performance on traditional S/390 workloads.

Each PU utilizes a superscalar design which provides all of the following:

- ▶ Decoding of 2 instructions per cycle
- ▶ Execution of 3 instructions per cycle (oldest must be a branch)
- ▶ In-order execution
- ▶ Out-of-order memory access

The *book* design, the centrally integrated crosspoint switch, and a large shared L2 cache that is logically and physically spread across the four books enables the movement of data from source to destination as efficiently as possible. The z9-109 server has further optimized the usage of cross-book memory operations. Each book has sixteen 2.7 GB/second STI links that form the base for high bandwidth I/O operations.

Other processor design features that contribute to the enhanced performance of z9-109 server processors are:

- ▶ The base clock rate is approximately .6 nanoseconds, compared to .83 nanoseconds in the base z990 machines.
- ▶ Floating-point performance is optimized for IEEE arithmetic (exploited by Java™ applications). Floating-point operations have now been fully pipelined at five stages for both IEEE (binary) and IBM floating point (hex) arithmetic. Additional optimization for BFP is present in z9-109 servers.
- ▶ A programmable Dynamic Address Translation (DAT) translator is used. This mechanism allows for later updates to the address translation algorithm (mapping virtual addresses to physical memory locations). A high-performance DAT function is complex and this design allows for corrections and design changes to the current algorithm, if necessary, as well as making the translator extendable for future architectural changes.
- ▶ A Second Level Translation Lookaside Buffer (TLB) is used. This mechanism provides a secondary cache for the DAT for both the instruction and data caches. The secondary buffers provide an additional 4K of dynamic address mapping over the current 512 entries available.
- ▶ Several cryptographic instructions have been added (and extended for the z9-109 server). These instructions for symmetric encryption/decryption are used to speed up SSL transactions, VPN data transfer, and data repository applications.
- ▶ A design for asymmetric mirroring for error detection was implemented with the zSeries servers. This provides a mechanism by which the mirrored execution used for error detection occurs 1 cycle delayed from actual operation. This allows the operation execution logic and circuitry to be optimized for performance.

The clock speed in the z9-109 server is faster than the clock speed of zSeries servers, of course, but the clock speed ratios are not a necessary indicator of performance differences. Improvements in logic design, logic element implementation, superscalar functions, various cache and TLB functions, and similar details may produce performance improvements beyond that indicated by clock ratios.

PR/SM performance factors

The operation of the internal PR/SM code has many effects on performance. Factors for the z9-109 server include:

- ▶ The server attempts to position all the real memory for an LPAR within one book. This is not always possible. No portions of real memory are lost due to this attempt.
- ▶ The server attempts to dispatch a CP (for an LPAR) that is in the same book as the real memory for that LPAR. If a CP from that book is not free, a CP from another book is used. A potential LPAR CP dispatch is not delayed waiting for a free CP in the target book.
- ▶ An attempt is made to redispach a logical CP to the same physical CP (even if there has been an intermediate user of that CP). This may allow some of the cache or TLB data to be reused.²¹ However, if that CP is not available, another CP will be used.

The effect of these functions is minor and is typically in the “noise level” of performance measurements when practical workloads are used. In practice, LPAR real memory may be split across books and CPs may be dispatched from any book. The L1 and L2 cache mechanism is so effective that the location of real memory (in the local book or in another book) is usually insignificant.

²¹ Minor changes to the z/OS dispatcher were needed for this function.

3.14 Program-directed re-IPL

A new function, program-directed re-IPL, allows an operating system to re-IPL itself. This function is supported for standard IPLs (based on CCWs via ESCON or FICON channels) and for SCSI IPLs (via FCP channels). A new SCLP call is provided with the z9-109 server to cause a re-IPL of an LPAR using the same IPL parameters as were used for the last manual IPL.

This function could be used by any operating system, but is primarily intended for Linux usage.

3.15 Open exchanges for FCP and FICON

An *open exchange* is part of FICON (and FC) terminology. An open exchange represents an I/O operation in progress over the channel. Many I/O operations can be in progress over these channels at any one time. For example, a disk I/O operation might temporarily disconnect from the channel while performing a seek operation or while waiting for a disk rotation. During this disconnect time, other I/O operations can be managed. In addition, these channels can multiplex data transfer for several devices at the same time. zSeries servers could sustain up to 32 such open exchanges per FICON (or FCP channel); the z9-109 server has increased this number to 64.

This larger number contributes to system performance. If the open exchange limit is reached, new I/O operations are refused by the channel. This results in necessary queuing and retries by the operating system.

3.16 CHPID mapping

The CHPID Mapping Tool is available on the IBM Resource Link™. The z9-109 server does not have CHPIDs assigned to I/O ports as part of the initial IBM setup process. CHPID numbers may be arbitrarily assigned to I/O ports; the I/O ports are identified by fixed PCHID numbers. In trivial cases, this assignment is easily done manually. In most practical cases, it can be quite complex and the use of the CHPID Mapping Tool is strongly recommended.

The CHPID Mapping Tool and an example of its use during migration from earlier servers is well described in *IBM eServer zSeries 990 Technical Introduction*, SG24-6863. The example is detailed and we have not copied the material into this document. We suggest the material in the z990 document is required reading for anyone undertaking a migration from an earlier server to a z9-109 server.

In its normal usage, the CHPID Mapping Tool uses the IBM-provided configuration file for *your* z9-109 server (before it is delivered) and your IOCP file from your current (older) server. The tool may require several iterations (with manual input) to resolve assignment conflicts and produce the best configuration for availability and performance. The result is a new IOCP file.

The CHPID Mapping Tool produces different results for a z9-109 server than for z990 or z890 servers. This is because the redundant I/O interconnect function is best utilized by having domain pairs in I/O cages connected to two different books whenever possible. This consideration did not exist in earlier servers.

3.17 LPARs

The z9-109 server can have up to 60 LPARs. The key element here is that the previous architectural limit of 15 (which used a 4-bit field in various LIC and software control blocks) has been extended. The zArchitecture functions use a one-byte field. There are a few additional changes that may be important:

- ▶ A z9-109 server runs only in LPAR mode. Basic mode (without LPARs) no longer exists.
- ▶ A maximum of 15 LPARs may be associated with a single channel subsystem.
- ▶ Partition identifiers, names, and numbers are a little more complex than with older servers, as explained in this section.

There are several identifiers associated with an LPAR:

- ▶ **LPAR identifier.** This is a number in the range 0 - X'3F' and is assigned by the user when defining LPAR Image profiles through the SE or HMC. It is unique across the z9-109 server. This identifier is returned by an STIDP or STSI instruction. This identifier is also known as the user logical partition id (UPID).
- ▶ **MIF Image id.**²² This number is defined through HCD or IOCP and is the partition number defined in the RESOURCE statement in an IOCP source statement. It is in the range 1 to X'F' and is unique *within* an LCSS. It does not need to be unique within a z9-109 server. This id is also known as the IID.
- ▶ **LPAR name.** This name is defined through HCD or an IOCP and is the partition name in the RESOURCE statement in an IOCP source statement deck. An LPAR name must be unique across the z9-109 server.

The LPAR identifier is used when defining activation profiles. The HCD field for the MIF Image has the label "Partition Number" and this is equivalent to the MIF Image id used in IOCP RESOURCE statements. We suggest you establish a naming convention for the LPAR identifier. For example, you could use the LCSS number concatenated to the MIF id. LPAR id 1A could mean LCSS 1, MIF id A. This fits within the allowed range of LPAR ids and conveys useful information to the user.

3.18 Hardware Management Console and Support Element

The Hardware Management Console (HMC) and Support Element (SE) were reorganized and have a new design on z9-109 servers. OS/2 is no longer used as a base operating system, but the HMC application has the same functionality as its predecessors.

HMC features

The HMC application can be accessed remotely from any standard Web browser. The major difference with the previous HMC Web interface application is that by default, all HMC functions are available to the user with the same *look and feel* characteristics of a local HMC desktop.

The first screen that shows up when the HMC application is accessed is no longer the *login* screen but the new *welcome screen* (see Figure 3-9 on page 46). There is an overall system status indicator, a hardware and operating system messages icon and a link, on the top, that drives to the login screen.

²² MIF means Multiple Image Facility. It was formerly known as EMIF for ESCON Multiple Image Facility because it was introduced for ESCON channels. The function is more general now, so the "E" portion of the name has been dropped.

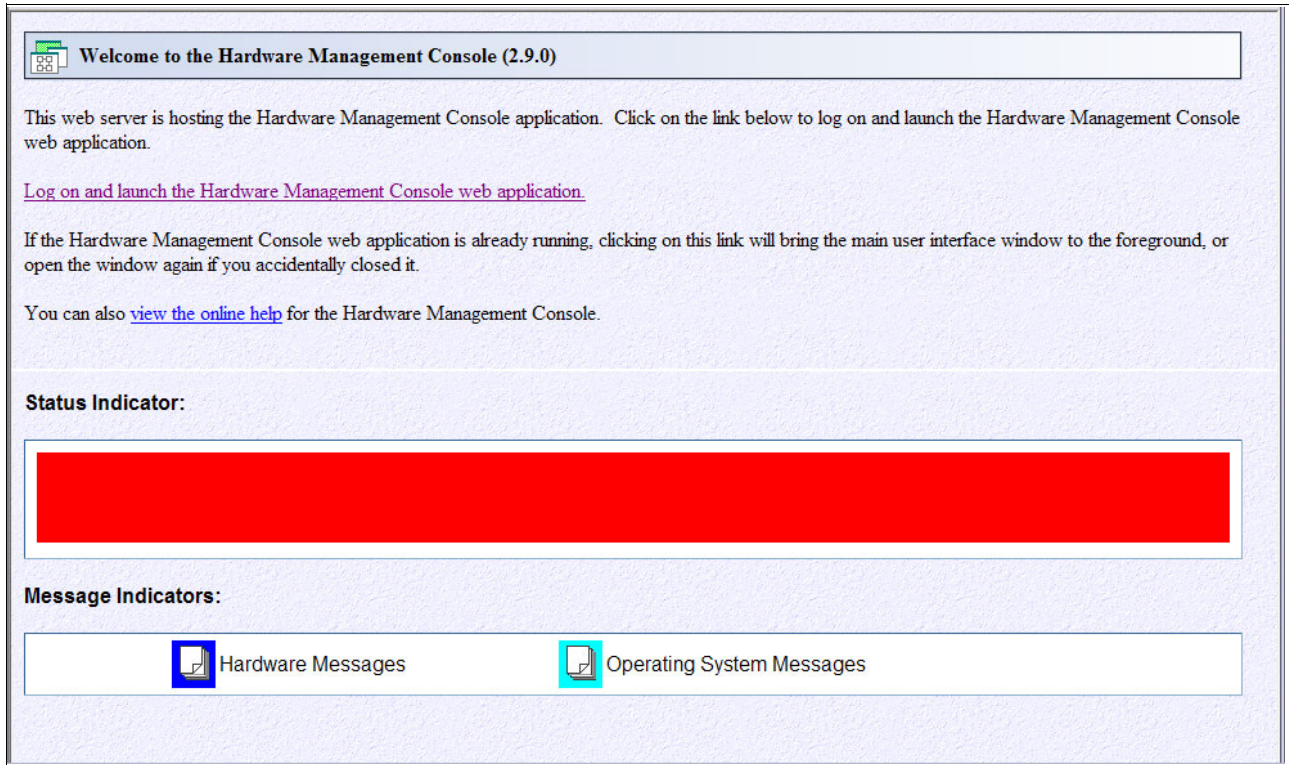


Figure 3-9 HMC welcome screen

The user interface has two user selectable styles: the *traditional* HMC style and the *tree Web-based* user style.

The traditional HMC style has the same format of the current HMC application and the look and feel is similar, but there are minor differences. For example, a drag and drop operation is replaced with a *select* operation. Figure 3-10 on page 47 provides an example of the traditional HMC style.

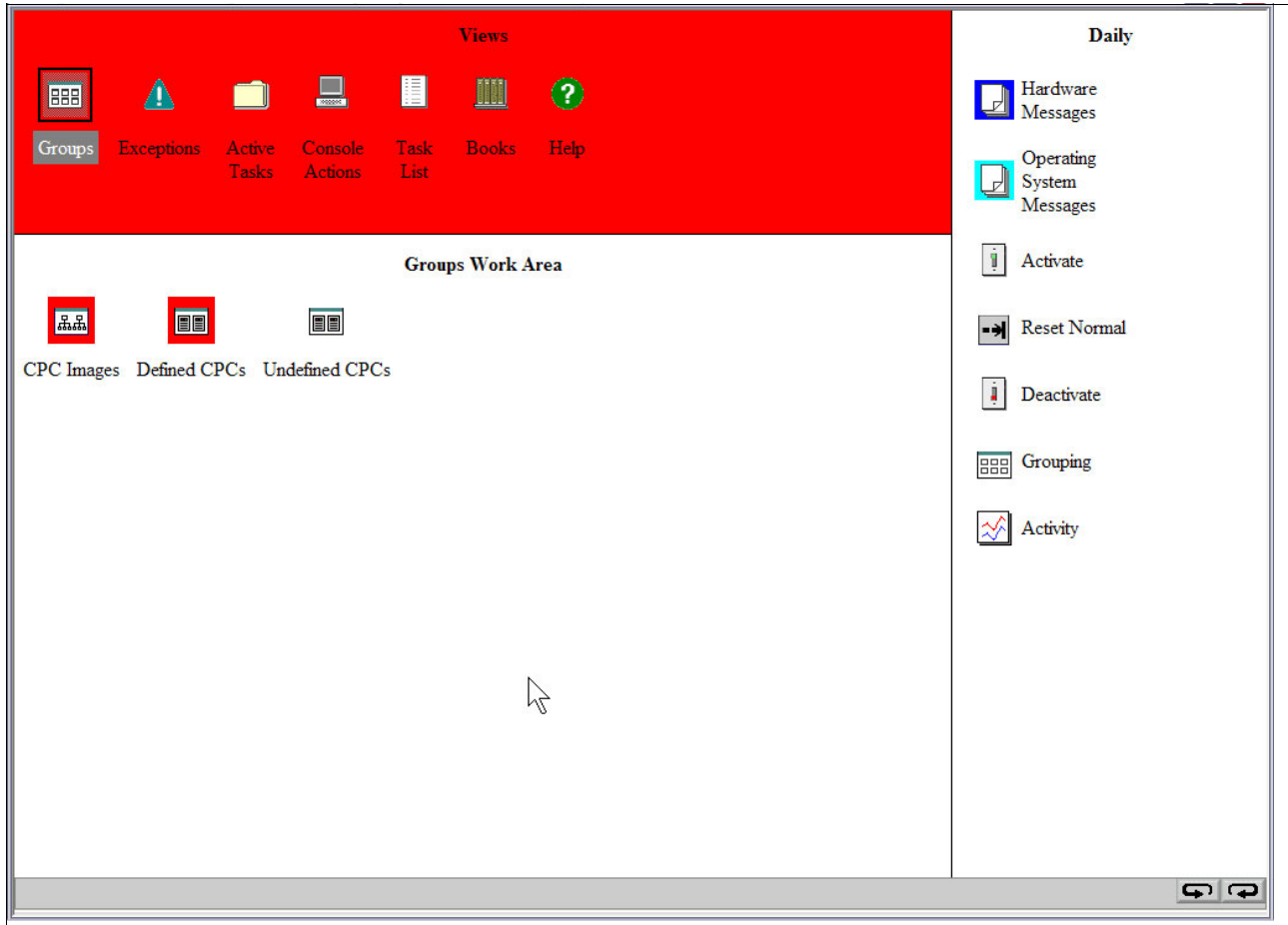


Figure 3-10 Traditional HMC style

The tree Web-based user style has a new design and is intended to be a standard graphic user interface style across all IBM eServer™ platforms. Figure 3-11 on page 48 shows a sample of this style. Note that there are three distinct areas: the *navigation area* on upper left, the *status and summary area* on the bottom left, and the *work area* on the right. The same tasks can be performed using either style.

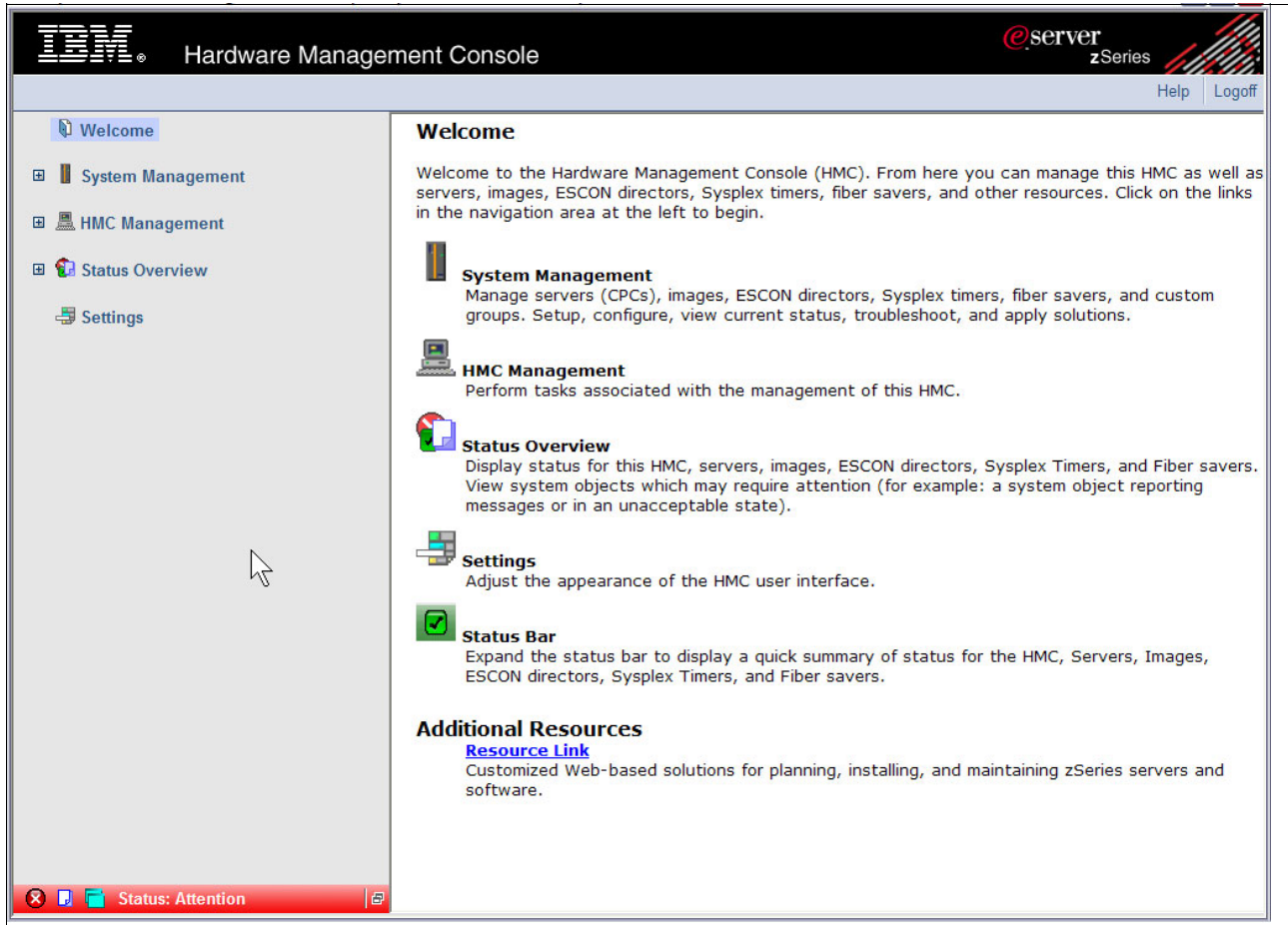


Figure 3-11 HMC tree Web-based user style

Figure 3-12 on page 49 shows a sample access to the IBM Resource Link site straight from the HMC console application.

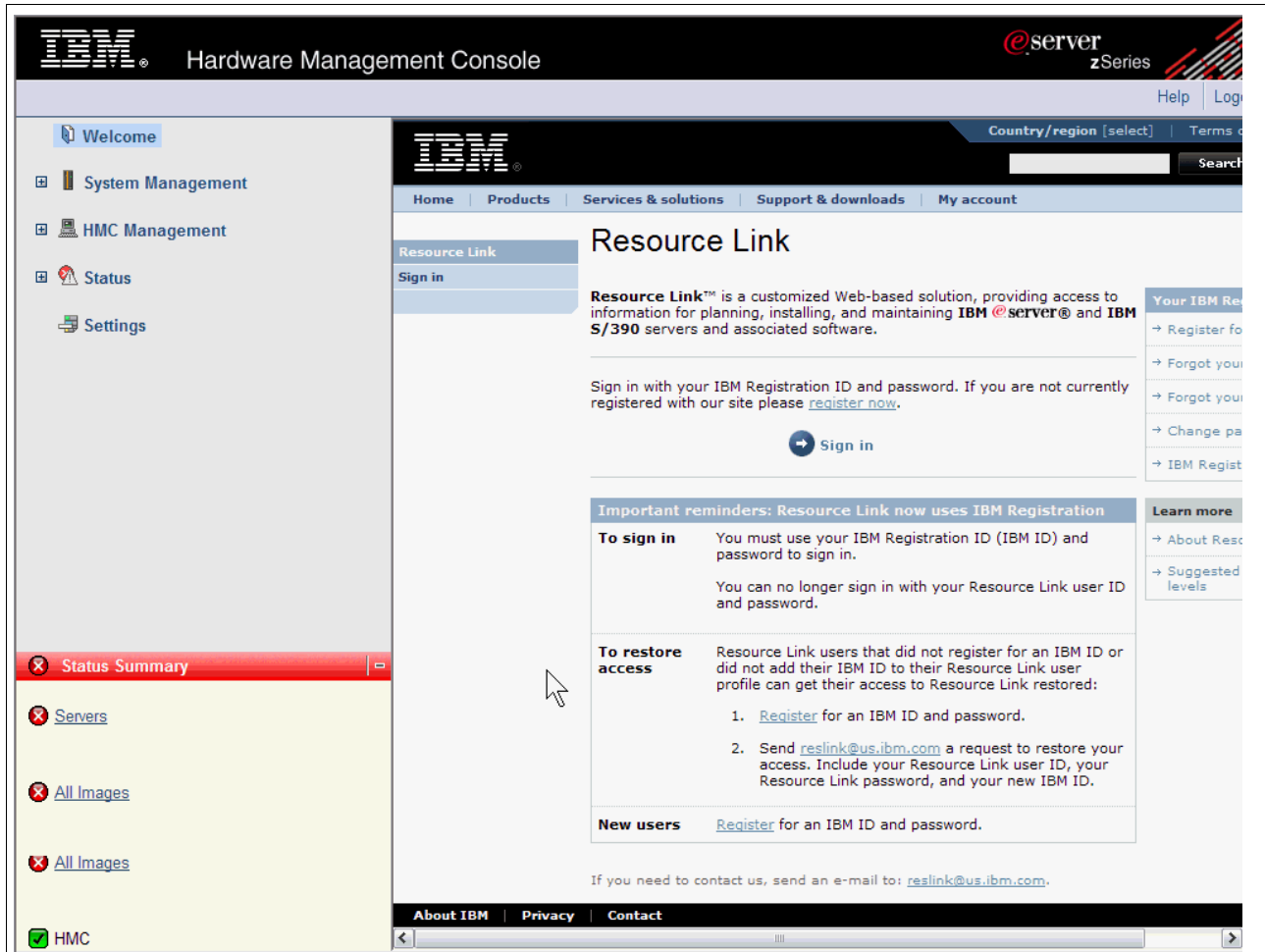


Figure 3-12 Resource link access through the HMC

The new *disconnect* option, provided with the logoff screen, will keep the session in the background until the next reconnect as long as the HMC or SE application is not restarted. When the same user ID tries to log on to the disconnected session, a *disconnect screen* is presented showing the options of reconnect, create a new session, or delete the old disconnected session. Figure 3-13 shows a sample of the logoff screen and the reconnect screen.

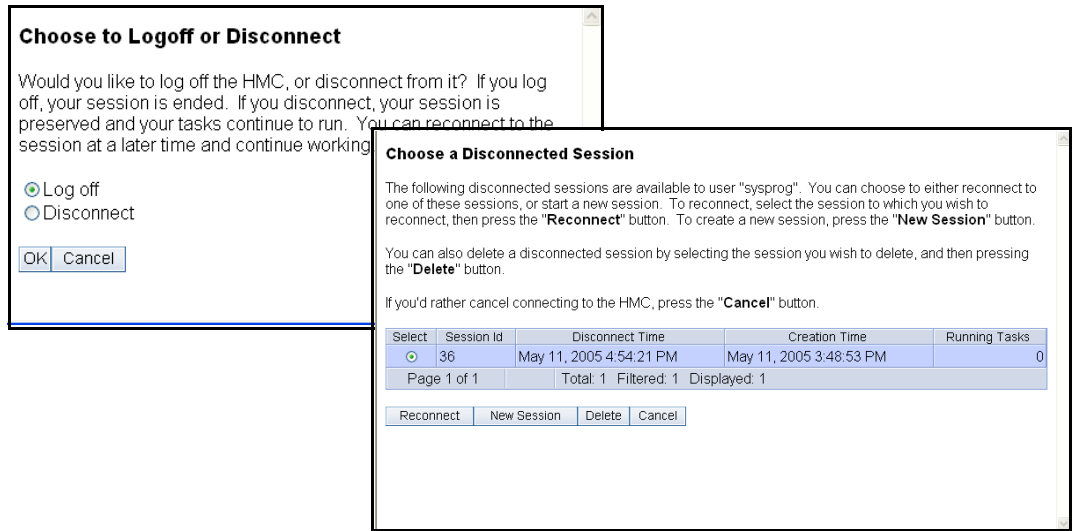


Figure 3-13 HMC logoff and reconnect screen

Security has been enhanced as expected for a critical Web-based application like the HMC application. Any communication with the HMC is certified based on SSL 128-bit encryption. Additionally, a dynamic and integrated firewall controls and limits all communications and connectivity to the HMC application.

The Remote Support Facility can be either using a modem (asynchronous communication is still supported) or through the LAN (more precisely, the Internet).

The new HMC application can coexist with previous HMC applications with the following minimum driver levels:

- ▶ 9672 G5/G6 (Driver 26)
- ▶ z800 and z900 (Driver 3G)
- ▶ z890 and z990 (Driver 55)

Moreover, the new HMC application can be installed on the current HMC desktop once the minimum hardware requisites are satisfied. See 5.1, “Hardware Management Console and Support Element” on page 62 for more information on HMC/SE, coexistence, and restrictions.



Software support

This chapter describes the operating system requirements and support considerations for the z9-109 system and its new features. It covers z/OS, z/VM, z/VSE, VSE/ESA™, z/TPF, TPF, and Linux on System z9. This information is subject to change; therefore, for more current information refer to the Preventive Service Planning (PSP) bucket (2094DEVICE).

Note that use of the following functions, new with the z9-109 system, *does not* extend to zSeries servers:

- ▶ 60 logical partitions
- ▶ Single system image
- ▶ 63.75K subchannels
- ▶ Multiple subchannel sets (MSSs)
- ▶ MIDAW facility
- ▶ N_Port ID virtualization
- ▶ FICON link incident report
- ▶ Program-directed re-IPL
- ▶ OSA-Express2 1000BASE-T Ethernet
- ▶ OSA-Express2 GARP VLAN Registration Protocol
- ▶ OSA-Express2 OSN support
- ▶ HiperSockets IPv6
- ▶ Crypto Express2
- ▶ Enhanced performance assists for z/VM Linux guests

4.1 Operating system support

Table 4-1 summarizes the minimum levels of operating systems required by the z9-109 system. z/OS V1 R4 must have either the Compatibility support feature or the Exploitation support feature installed to run on the z9-109.

Table 4-1 z9-109 system operating system support summary

Operating System	ESA/390 (31-bit Mode)	z/Architect (64-bit mode)	Support
z/OS V1 R4	Yes ^a	Yes	Basic support
z/OS V1 R5	No	Yes	Basic support
z/OS V1 R6	No	Yes	
z/OS V1 R7	No	Yes	Support for new HW functions
z/VM V4 R4	Yes	Yes	Basic support
z/VM V5 R1	No	Yes	Basic support
z/VM V5 R2	No	Yes	
Linux on System z9 - 64 bit	No	Yes	
Linux 390 - 31 bit	Yes	No	
VSE/ESA V2 R6	Yes	No	Basic support
VSE/ESA V2 R7	Yes	No	Basic support
z/VSE V3 R1	Yes	No	Basic support
TFP V4 R1	Yes	Yes	
z/TFP V1 R1	Yes	Yes	

a. Support through the IBM Accommodation Offering only available for z/OS 1.4

Basic support means that the operating system is capable of recognizing the z9-109 hardware features, but does not exploit them. *Support for new HW* means that *some* z9-109 hardware features are exploited. The features exploited may depend on a particular operating system. For example, the program-directed re-IPL feature will be exploited only by Linux on System z9.

When running in 64-bit mode (z/Architecture), Expanded Storage (ES) is supported by z/VM, but not by z/OS. However, if z/OS is used in 31-bit mode (ESA/390), then ES may be used.

The remainder of this chapter examines specific features of the z9-109 server and indicates the appropriate operating system levels needed for that feature. If an operating system is not listed for a feature or the information is blank it means that we were unable to determine the status at the time of writing. *In all cases, various PTFs may be needed with the operating system level indicated.*

4.2 60 logical partitions

This feature, which is not present on any earlier zSeries servers, allows the system to be configured with up to 60 LPARs. (There is a limitation of 15 LPARs per logical channel subsystem; it is necessary to configure four logical channel subsystems to reach 60 LPARs.)

Table 4-2 60 logical partitions software support

Operating System	Support Requirements
z/OS V1 R7	Included
z/OS V1 R5, R6	PTFs required
z/OS V1 R4	Exploitation support feature plus PTFs
z/VM V5 R2	Included
z/VM V5 R1	PTFs (for VMHCD)
z/VM V4 R4	PTFs (for VMHCD)
z/VSE V3 R1	PTFs
z/TPF V1 R1	
TPF V4 R1	
Linux on System z9	SuSE SLES 8 and later; RedHat RHEL 3 and later

4.3 Single system image

A single system image (z/OS in an LPAR, for example) can control several processors (CPs, zAAPs, ILFs, as appropriate). Table 4-3 represents the maximum number of processors supported by each system image.

Table 4-3 Single system image software support

Operating System	Total CPs+zAAPs ^a / IFLs supported in a single system image
z/OS V1 R7	Up to 32
z/OS V1 R6	Up to 32
z/OS V1 R4, R5	Up to 16
z/VM V5 R1, R2	Up to 24
z/TPF V1 R1	Up to 32

a. Total CPs and zAAPs refers to the sum of both these PU characterizations.

4.4 63.75K subchannels

Systems prior to the z9-109 system reserved 1024 subchannels out of the potential maximum of 64K subchannels. The z9-109 system has reduced this reserved number to 256 subchannels, thus increasing the number of subchannels available. (The reserved subchannels exist only in subchannel set 0; no subchannels are reserved in subchannel set 1.) The informal name, 63.75K, represents $63 \times 1024 + .75 \times 1024$. Remember that *included* may mean that PTFs are required.

Table 4-4 63.75K subchannels software support

Operating System	Support Requirements
z/OS V1 R7	Included
z/OS V1 R6	Included
z/OS V1 R5	Included
z/OS V1 R4	Exploitation support feature required
z/VM V5 R2	Included
z/VM V5 R1	PTFs (for HCD)
z/VM V4 R4	PTFs (for HCD)
z/TPF V1 R1	
TPF V4 R1	

4.5 Multiple subchannel sets (MSS)

Multiple subchannel sets are new with the z9-109 system. They provide a mechanism for addressing more than 63.75K I/O devices (or aliases).

Table 4-5 MSS software support

Operating System	Support Requirements
z/OS V1 R7	Included
z/OS V1 R5, R6	Coexistence - requires SPE ^a for HCD and IOCP
z/OS V1 R4	Coexistence - requires Exploitation support feature plus SPE for HCD and IOCP
z/VM V5 R2	Not supported
z/VM V5 R1	Not supported
z/VM V4 R4	Not supported

a. SPE stands for small program enhancements.

The following SPEs should be applied for z/OS operating systems earlier than V1 R7:

- ▶ OA08197 - IODF support
- ▶ OA07875 - HCD support

Multiple subchannel sets are not supported when z/OS is running as a guest under z/VM.

4.6 MIDAW facility

The MIDAW facility is new with the z9-109 system; it provides a more efficient replacement for data-chained CCWs.

Table 4-6 MIDAW facility software support

Operating System	Support Requirements
z/OS V1 R7	Included
z/OS V1 R6	PTFs - planned available date in 4Q2005
z/VM V5 R2	No support
z/VM V5 R1	No support
z/VM V4 R4	No support

The MIDAW facility is not supported when running as a guest under z/VM.

4.7 N_Port ID virtualization

N_Port ID virtualization is new with the z9-109 system; it provides a way to allow multiple system images (in LPARs and/or z/VM guests) to use a single FCP channel as if each were the sole user of the channel. Note that this feature may be used with earlier FICON features that have been carried forward from earlier machines.

Table 4-7 N_Port ID virtualization software support

Operating System	Support Requirements
z/VM V5 R1 , R2	PTFs - available on 1H2006
z/VM V4 R4	PTFs - available on 1H2006
z/VSE V3 R1	PTFs
Linux on System z9	Support will be delivered through IBM and distribution partners. For more information go to: http://www.ibm.com/developerworks/linux/linux390/

4.8 FICON link incident report

FICON link incident reporting is designed to allow an operating system image (without operator intervention) to register for link incident reports.

Table 4-8 Link incident report software support

Operating System	Support Requirements
z/OS V1 R7	Included

4.9 Program-directed re-IPL

Program-directed re-IPL is designed to allow an operating system on a z9-109 server to re-IPL without operator intervention. This function is supported for both SCSI and ECKD™ devices.

Table 4-9 Program-directed re-IPL software support

Operating System	Support Requirements
Linux on System z9	Support will be delivered through IBM and distribution partners. For more information go to: http://www.ibm.com/developerworks/linux/linux390/

4.10 OSA-Express2 1000BASE-T Ethernet

This adapter has multiple interfaces, with QDIO (CHPID type OSD), non-QDIO (CHPID type OSE), and local 3270 emulation (CHPID type OSC).

Table 4-10 OSA-Express2 1000BASE-T software support

Operating System	CPHID type OSC	CHPID type OSD	CHPID type OSE
z/OS V1 R7	Included	Included	Included
z/OS V1 R6	Included	Included	Included
z/OS V1 R5	Included	Included	Included
z/OS V1 R4	Exploitation support feature	Exploitation support feature	Exploitation support feature
z/VM V5 R2	Included	Included	Included
z/VM V5 R1	Included	Included	Included
z/VM V4 R4	PTFs	Included	Included
z/VSE R3 V1	Included	Included	Included
VSE/ESA V2 R7	Included	Included	Included
z/TPF V1 R1		Included	
TFP V4 R1		PUT 13 plus PTFs	
Linux on System z9		SuSE SLES 8, SLES 9 Red Hat RHEL 3	

4.11 OSA-Express2 GARP VLAN Registration Protocol (GVRP)

GVRP support allows an OSA-Express2 port to register or de-register its VLAN IDs with a GVRP-capable switch and dynamically update its table as the VLANs change.

Table 4-11 GVRP software support

Operating System	Support requirements
z/OS V1 R7	Included
z/VM V5 R2	Included

Operating System	Support requirements
Linux on System z9	Support will be delivered through IBM and distribution partners. For more information go to: http://www.ibm.com/developerworks/linux/linux390/

4.12 OSA-Express2 OSN support

Channel Data Link Control (CDLC), when used with the Communication Controller for Linux, emulates selected functions of IBM 3745/NCP operations. The port used with the OSN support appears as an ESCON channel to the operating system. This support may be used with any OSA-Express2 feature, except the 10 GbE LR feature.

Table 4-12 OSN software support

Operating System	CHPID type OSN
z/OS V1 R7	Included
z/OS V1 R6	PTFs
z/OS V1 R5	PTFs
z/OS V1 R4	Exploitation support feature plus PTFs
z/VM V5 R2	Included
z/VM V5 R1	PTFs
z/VSE V3 R1	PTFs
z/TPF V1 R1	Included
TPF V4 R1	Included
Linux on System z9	Support will be delivered through IBM and distribution partners. For more information go to: http://www.ibm.com/developerworks/linux/linux390/

4.13 HiperSockets IPv6

IPv6 is expected to be a key element in future networking. The IPv6 support for HiperSockets permits compatible implementations between external networks and internal HiperSocket networks.

Table 4-13 HiperSockets IPv6 software support

Operating System	Support requirements
z/OS V1 R7	Included
z/VM V5 R2	PTFs - available on 1H2006
Linux on System z9	Support will be delivered through IBM and distribution partners. For more information go to: http://www.ibm.com/developerworks/linux/linux390/

4.14 Crypto Express2

Most cryptographic usage is through ICSF, which optimizes use of whatever hardware cryptographic functions are available.

4.14.1 PCI-X support

The Crypto Express2 feature provides two PCI-X adapters. On the z9-109 server, these adapters can be configured as either secure processors or accelerators, or one of each. Web-deliverable support for z/OS V1 R6 and z/OS V1 R7 can be found at:

<http://www.ibm.com/eserver/zseries/zos/downloads>

z/VSE, and Linux on System z9 support clear key RSA operations only. z/VM V5 R1, and later releases, support clear and secure key operations.

Table 4-14 Crypto Express2 software support

Operating System	Support requirements
z/OS V1 R7	Web deliverable support
z/OS V1 R6	Web deliverable support
z/OS V1 R4, R5	PTFs
z/VM V5 R2	PTFs - planned date 12/16/2005
z/VM V5 R1	PTFs - planned date 09/16/2005
z/VSE V3 R1	PTFs - availability 1Q2006
VSE/ESA V2 R7	PTFs - availability 1Q2006 (IBM TCP/IP for SSL support)
Linux on System z9	Support will be delivered through IBM and distribution partners. For more information go to: http://www.ibm.com/developerworks/linux/linux390/

4.14.2 CPACF enhancements

CP assist for cryptography function (CPACF) has additional functions available through the cryptographic instructions present in all z9-109 system PUs.

Table 4-15 CPACF software support

Operating System	Support requirements
z/OS V1 R7	Web-deliverable support
z/OS V1 R6	Web-deliverable support
z/VM V5 R1, R2	PTFs
z/VM V4 R4	PTFs
Linux on System z9	Support will be delivered through IBM and distribution partners. For more information go to: http://www.ibm.com/developerworks/linux/linux390/

4.15 Enhanced performance assists for z/VM Linux guests

Linux on System z9 images running as guests under z/VM can take advantage of the performance assist enhancements.

Table 4-16 Enhanced performance assists for z/VM guests software support

Operating System	Support requirements
z/VM V5 R2	Included
Linux on System z9	Support will be delivered through IBM and distribution partners. For more information go to: http://www.ibm.com/developerworks/linux/linux390/

4.16 Support summary

The following summary indicates the *lowest* levels of various operating systems that support the indicated functions. These lower levels may require PTFs, SPEs, or other additional software to provide the function. z/OS 1.4 requires either the Compatibility support feature or the Exploitation support feature to run on a z9-109 server.

Table 4-17 Operating system support summary

Function	z/OS	z/VM	z/VSE VSE	z/TPF TPF	Linux on System z9
Base z9-109 support	1.4	4.4	3.1 2.7	1.1 4.1	SLES 8 RHEL 3
60 logical partitions	1.4 ^a	4.4	3.1 N/A	1.1 4.1	SLES 8 RHEL 3
63.75K subchannels	1.4	4.4			
Multiple subchannel sets	1.7				
MIDAW facility	1.6				
N_Port ID virtualization		5.1	3.1 N/A		TBD ^b
FICON Link incident report	1.7				
Program directed re-IPL					TBD ^b
OSA Express2 1000BaseT (CHPID OSC)	1.4 ^a	4.4	3.1 2.7		
OSA Express2 1000BaseT (CHPID OSD)	1.4 ^a	4.4	3.1 2.7	1.1 4.4	SLES 8 RHEL 3
OSA Express2 1000BaseT (CHPID OSE)	1.4 ^a	4.4	3.1 2.7		
OSA Express2 OSN	1.4 ^a	5.1	3.1 N/A		TBD ^b
OSA Express2 GVRP	1.7	5.2	N/A	NA	TBD ^b
Hipersockets IPv6	1.6	5.2			

Function	z/OS	z/VM	z/VSE VSE	z/TPF TPF	Linux on System z9
CryptoExpress2	1.6	5.1	3.1 2.7		TBD ^b
CPACF Enhancements	1.6	4.4			TBD ^b
Enhanced performance assist for z/VM guests		5.2			TBD ^b
Server Time Protocol ^c	1.7				

a. Exploitation support feature is required.

b. IBM is working with Linux Distribution Partners to provide this function.

c. Statement of Direction.

For the most current operating system maintenance information go to the following Web sites:

► z/OS

<http://www-1.ibm.com/servers/eserver/support/zseries/zos/>

► z/VM

<http://www-1.ibm.com/servers/eserver/support/zseries/zvm/planning.html>

► z/TPF

<http://www-306.ibm.com/software/http/tpf/pages/maint.htm>

► z/VSE

<http://www-1.ibm.com/servers/eserver/zseries/zvse/support/preventive.html>

► Linux on System z9

<http://www-1.ibm.com/servers/eserver/zseries/os/linux/support.html>



General considerations

This chapter discusses several areas that should be considered when moving from an earlier zSeries server (z900 or z990) to a z9-109 system. This includes:

- ▶ Hardware Management Console and Support Element
- ▶ Cabling services and cable planning
- ▶ I/O channel considerations
- ▶ I/O configuration definition and management
- ▶ Capacity upgrades
- ▶ Logical planning
- ▶ Coupling Facility planning
- ▶ Operating system migration

5.1 Hardware Management Console and Support Element

The Hardware Management Console (HMC) delivered with the z9-109 server includes a completely new implementation of the HMC application code. It is no longer based on OS/2 and is now a *closed system* in which the underlying operating system is not accessible by the user. Additional applications cannot be installed. The HMC application itself is the only possible application.

Functionally, the HMC has not changed significantly. It still is the hardware control point with the same functions provided by earlier HMC implementations. It includes functions introduced in recent releases, such as the integrated 3270 and ASCII consoles. There are now two user interface options:

- ▶ A traditional interface that is generally similar to what is seen on previous HMC consoles.
- ▶ A new *tree interface* that is oriented to more hierarchical use.

The appearance of the two interfaces is quite different, but the resulting functionality is the same. Both interfaces may be used at the same time when multiple HMCs are involved.

The mechanism for viewing and using the interfaces has changed. There is no direct access to the interfaces on the HMC itself; a browser must be used. The appearance, functionality, and performance is the same whether the browser runs on the HMC or another PC. (This statement assumes reasonable LAN performance between the HMC and PC.) The HMC implementation includes an HTTP *server* as its user interface; the user connects to this interface with a browser.

Note that this is quite different from earlier HMC functions that allowed browser connections. These were either very limited in functionality (when using a normal browser interface) or involved a complete remote desktop image (using Desktop on Call or a similar interface). The remote desktop image required much more intensive LAN usage and often had somewhat sluggish performance. The full browser interface of the new implementation requires only “normal” LAN performance.

At least one HMC is needed to run a z9-109 system. Several HMC operators, at different locations, can all use the same HMC simply by accessing it through a browser. Multiple HMCs might be desired for redundancy.

The new HMCs provided with z9-109 systems have only Ethernet LAN connections. Token ring adapters are not available. However, older HMCs that are upgraded to run the new HMC code may have token ring adapters and these are still supported.

Mixed HMC levels

The new HMC level may be used to control z9-109 systems, zSeries servers, and G5 and G6 systems. Support element patches may be needed on these earlier machines. Systems prior to G5 cannot be accessed by the new HMC application. The driver levels required on these systems (if they are to be accessed through the new HCM level) are as follows:

Product	Driver
zSeries z990, z890	driver 55
zSeries z900, z800	driver 3G
G5 or G6	driver 26

The new HMC software can be installed on earlier HMC units with 1GB memory and a 40 GB hard disk (feature codes 0075 or 0076). These units can use token ring adapters they may already have installed.

Older HMCs (earlier drivers) cannot be used to operate z9-109 systems.

5.1.1 Support elements

Every z9-109 system has two identical support elements (SEs). These are IBM Thinkpads located under the system covers in the Z frame, one on top of the other. The first SE, on the top, is the primary; the other, on the bottom, is used for backup purposes. The SE is intended to be used only by IBM Service Representatives. However, it may be used to manage the system in the event of an HMC failure.

Support elements on z9-109 systems have 1GB of memory, a 60 GB hard disk, no diskette drive (internal or external), and a CD/DVD ROM drive. Two Ethernet adapters are provided; token ring is not available.

5.1.2 HMC and SE connectivity

Configuring and ordering SEs and HMCs requires an understanding of the LAN interfaces involved. Both units normally have two LAN interfaces. The general rules are:

- ▶ An SE is connected to a given HMC by one or two LANs. In simple situations, the second LAN interfaces included with SEs and HMCs are not used. If a second LAN is connected it will be used for automatic failover.
- ▶ The two LAN interfaces on an SE may be used to connect to one or more HMCs, using two independent LANs.
- ▶ Each SE on a z9-109 system has two Ethernet adapters that may be used for HMC connectivity.
- ▶ A default HMC configuration has two Ethernet interfaces. However, token ring interfaces can exist if the current HMC hardware was upgraded to the new HMC driver level and already had a token ring card installed.
- ▶ The HMC desktop must have at least one Ethernet interface. There is no other way to communicate with the z9-109 SE.

A very common arrangement is to use at least two HMCs, one near the system and the other near the general operations control area. With the new HMC design this may no longer be necessary because a browser (based in an inexpensive PC) can be used at one location to provide a second interface to the HMC in the other location.

You can purchase any reasonable number of HMCs to use with the system.²³ You have the following elements for planning your HMC and SE installation on your system:

- ▶ Four Ethernet cables (each 50 feet (about 15m) long) are provided with the z9-109 system. Potentially, these connect to the two Ethernet adapters in each SE. In practice, only one of the Ethernet connections in each SE is normally used.
- ▶ A new HMC with two Ethernet features. An existing HMC might have one Ethernet and one token ring adapter.
- ▶ A display for the HMC.
- ▶ An Ethernet switch. This is ordered automatically.
- ▶ A modem interface (use is optional).

The SE and HMC Ethernet ports can run at 10 Mbps or 100 Mbps and auto-sense the LAN speed. Figure 5-1 illustrates a basic SE/HMC configuration.

²³ Up to 32 HMCs can be used to control your z9-109 systems.

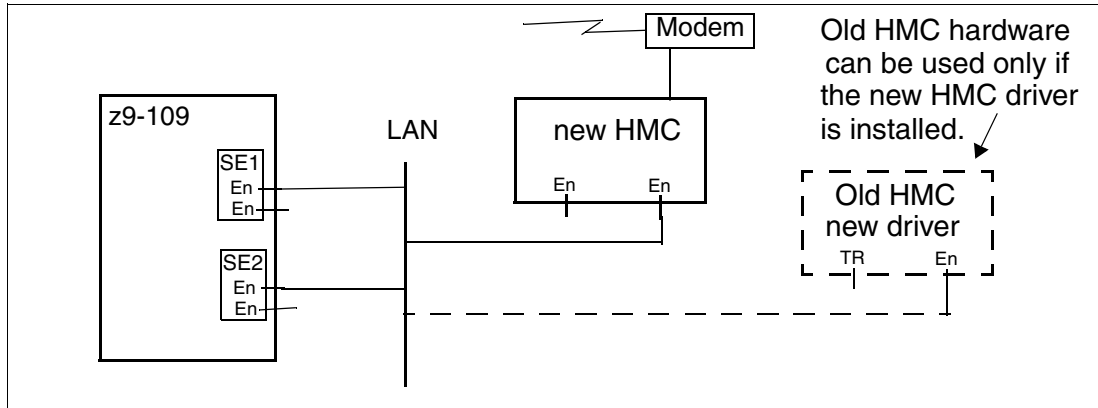


Figure 5-1 Basic SE - HMC connections

The modem shown in the sketch is required if a LAN connection from the HMC to the IBM Network is not available. It is used for Remote Support Facility (RSF) connections to transfer microcode updates (MCLs) and system status information via automated scheduled transmissions. Also, if a system fault occurs, the HMC places a call to the IBM Support System, alerting IBM and transferring error information and logs to reduce the impact of unplanned outages.

The figure contains only a single z9-109 system. SEs from other systems (z9-109 systems, zSeries, G5 or G6) could also be connected to the LAN and operated from the HMC. Other HMCs (new or at a compatible driver level) could be connected to the LAN. In this situation, all systems (assuming they have appropriate driver levels and patches) could be operated from the new HMCs, but the z9-109 systems could not be operated from the older HMCs.

HMC Version 2.9.0 or later does not use Distributed Console Access Facility (DCAF) to access other console screens. DCAF has been replaced by a version of Desktop-on-Call (DTC) suitable for the *closed system* HMC. Using DTC it is possible to access the desktops of other consoles, such as a Sysplex Timer, an ESCON Director, any zSeries server.

The only network protocol supported by the new HMC application is TCP/IP. Several specific considerations are important:

- ▶ The auto-discovery function using NetBios, which automatically located SEs and HMCs, is no longer present. The HMC now uses an IP Multicast mechanism instead of NetBios to automatically discover SEs and HMCs. If IP Multicast is disabled in the network then auto-discovery will not happen.
- ▶ SNA protocols are no longer supported. All SNA-based functions have been replaced by the IP equivalent. For example, tn3270 sessions are used instead of SNA 3270 sessions.
- ▶ The Remote Support Facility (RSF), also known as the *call home* function, can be done through the LAN (if the HMC desktop has access to the Internet) or via an asynchronous modem connection. SDLC modem connections are not supported.

SEs and HMCs may be connected through public LANs and this may be the best option for the *call home* function. The other option is the use of a dial-up asynchronous line. However, the amount of data traffic may, at times, create substantial delays when depending on a dial-up connection.

A browser connection to a new HMC always uses SSL encryption. A user public key certificate must be supplied with a new HMC installation. Every browser connection to an HMC requires a userid and a password. *Be certain to change the default userids and passwords as soon as possible.*

5.1.3 Planning recommendations

We suggest that a certain amount of planning and early actions are important for smooth implementation of the new HMCs. Important points to consider include:

- ▶ If possible, installation of a new HMC (or the new driver on an appropriate older HMC) should be done well before a z9-109 system is expected. Hands-on usage and training can be done while using the new HMC on existing zSeries or G5 and G6 machines.
- ▶ Hands-on usage and training for the new HMC is important, whether before a new z9-109 system arrives or afterward. The traditional HMC GUI provided on the new HMC is fairly close to that of older HMC versions but there are still differences. Anyone clearly understanding the functions being used should not have a problem, but operators attempting to follow existing *scripts* or *run books* may need some practice on the new HMC level.
- ▶ Anyone wanting to work with the new *tree* GUI may need some practice time.
- ▶ LAN IP assignments for the SEs and HMCs should be designed before a z9-109 system arrives.
- ▶ There are obviously security concerns if the SEs and HMCs will be placed on public LANs. The considerations are complex and beyond the scope of this introductory document. However, they should be addressed before the arrival of a z9-109 system.
- ▶ If the public LAN will be used, then connection provisions may be needed in corporate firewalls.
- ▶ An HMC can now use DHCP to obtain its IP addresses for its Ethernet adapters.

5.2 Cabling services and cable planning

Starting with the zSeries z800 server, fiber optic cables, cable planning, labeling, and installation became customer responsibilities for new installations and upgrades. *With a few exceptions, fiber cables cannot be ordered as part of your system order.*²⁴ Several vendors, including IBM, offer cables and cable-related services through separate contracts. *You must arrange a separate contract for fiber cables and their installation.*

Note that ICB-3 and ICB-4 cables are provided as part of a system order; these are not addressed in the optional contract services discussed here. Also, remember that parallel channels are not directly supported on the z9-109 system and parallel channel cables are not addressed by these contract services.

To better serve such cabling needs, IBM Networking Services offers a set of services. These services take into consideration the requirements for all of the protocols and media types supported on the z9-109 system and zSeries servers, and can encompass a data center, a storage area network (SAN), local area networks (LAN), or the end-to-end enterprise.

These services involve many details; the brief description here is intended only as an introduction. Your IBM representative can provide more information. The services are divided into two categories: basic cable services (with three options) and Enterprise fiber cabling services (with two options).

Briefly, these offerings include:

- ▶ Basic cable services to provide individual fiber optic cables (jumper cables, conversion kits, MCP cables) for connecting z9-109, z990, z890, z900, and z800 systems and

²⁴ Likewise, fiber optic conversion kits and mode conditioning patch cables are not orderable as part of a z9-109 system order.

servers. There are three options for this basic service (options 1 and 2 can be combined within one contract to provide complete upgrade coverage):

- Option 1 - *Fiber optic jumper cabling package*. This option includes planning, new cables, installation, and documentation. An analysis of the z9-109 system channel configuration, I/O devices, and any existing fiber optic cabling is required to determine the appropriate fiber optic cables that must be supplied.
 - Option 2 - *Fiber optic jumper cable migration and reuse for a zSeries upgrade*. This option includes planning, reuse of existing cables, and documentation. IBM organizes the existing fiber optic cables based upon the new zxxx connection details. Relabeling, rerouting, and reconnecting to the appropriate z9-109 system channels is performed. New cables are not offered as a part of this option.
 - Option 3 - *Fiber optic jumper cables and installation*. With this option, the customer does the planning and tells IBM what they need. The service includes new cables, installation, and documentation.
- Under the Enterprise Fiber Cabling Services umbrella there are two options to provide IBM Fiber Transport System (FTS) trunking commodities (fiber optic trunk cables, fiber harnesses, panel-mount boxes) for connecting to the z9-109, z990, z890, z900, and z800 systems and servers:
- Option 1 - *zSeries fiber optic trunk cabling package*. With this option IBM reduces the cable clutter under the floor. An analysis of the total IBM data center fiber channel configuration and any existing fiber optic cabling is performed to determine the required FTS fiber optic trunking commodities (trunk cables, harnesses, panel-mount boxes). This option includes planning, FTS fiber optic trunking commodities, installation, and documentation.
 - Option 2 - *Enterprise fiber cabling services*. With this option IBM organizes the entire enterprise. This option includes enterprise planning, new cables, fiber optic trunking commodities, installation, and documentation. This is the most comprehensive set of services.

Additional information is documented in the IBM Redbook *IBM eServer zSeries Connectivity Handbook*, SG24-5444.

Fiber cable connectors

Some fiber cable connectors used on older servers are not used on the z9-109 system. Figure 5-2 shows the connectors being used on z9-109 systems.

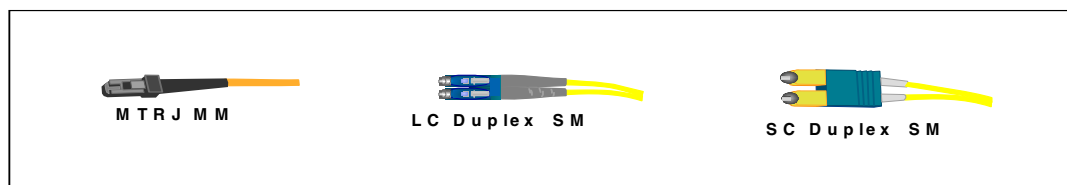


Figure 5-2 z9-109 system fiber connectors

Reuse of existing cables may require conversion kits. Figure 5-3 illustrates the most common conversion kits available through the IBM fiber cabling services offering.

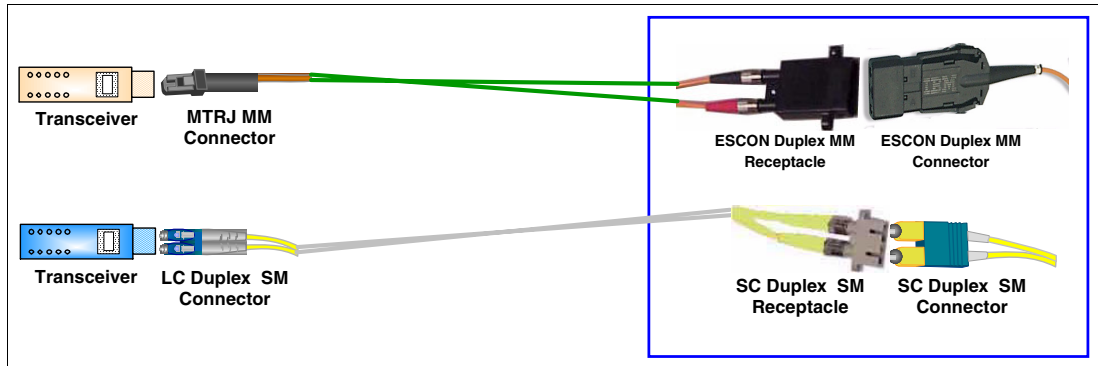


Figure 5-3 Under floor connector conversion kits

Reuse of a multimode trunk cable is also something to consider. If you have a multimode fiber infrastructure, the use of Mode Conditioning Patch (MCP) cables allows reuse of that infrastructure with LX features that normally attach to single-mode fiber connectors. Two MCP cables are required per link, that is, one at each end. There are some restrictions when using MCP:

- ▶ Restricted to 1-Gigabit links only
- ▶ Not supported with 2-Gigabit ISC-3 or FICON Express features
- ▶ Feature-to-device distance limited to 550 meters (1804 feet)

You should consider migrating from a multimode infrastructure to a single-mode infrastructure. More specific information about fiber cables may be found in *Planning for Fiber Optic Links*, GA23-0367.

5.3 I/O channel considerations

This section discusses topics that should be considered when dealing with the I/O channel options available with the z9-109 server.

For a list of the supported I/O features on the z9-109 server, refer to Appendix C, “Channel options” on page 83.

5.3.1 Parallel channel migration

Parallel channels²⁵ are not supported on z9-109 systems. If you still need parallel channel devices²⁶ you must use *ESCON converters* to connect the parallel devices to ESCON channels. IBM recommends the use of the Optica ESCON Converter. The IBM 9034 units may also be used, but these cannot be ordered through normal IBM ordering channels.

As a general recommendation on parallel channels:

- ▶ Keep only selected parallel devices and consolidate them onto as few channels as possible. Remember that *daisy chaining* is possible with ESCON converters.
- ▶ Consider replacing the following parallel channel devices:
 - IBM 3174 control units, used for 3270 operating system consoles and TSO/CICS 3270 terminals can be replaced by the OSA-Express ICC function.

²⁵ Parallel channels are often known as *bus and tag* or OEMI channels, named for the functions of the two cables used for these channels.

²⁶ We use the term *devices* here to include the associated control units. It is the control units, of course, that actually use the channels.

- Some Communication Control Units (CCU), such as 3745-type units running NCP, can be replaced by an OSA-Express2 1000BASE-T Ethernet (CHPID type OSN) and Communication Controller software for Linux on zSeries (CCL).

5.3.2 ESCON channel planning

On z9-109 systems, ESCON channels are always delivered with 16-channel I/O feature cards. The z900 4-port cards (FC2313) cannot be used. Up to 15 channels on each card are available for use; the last channel is reserved as a spare. In practice, any unallocated ports on the card can act as a spare.

A minimum of two ESCON channel feature cards are always installed if any ESCON channels are configured for the processor. ESCON channels are ordered in groups of four. For configurations greater than 28 ESCON channels, individual (not pairs) ESCON channel cards are added as necessary. The active channels are distributed across the physical cards to provide additional redundancy.

These ESCON cards, which are also used with z990, z890, z900, and z800 servers, use MT-RJ connectors. These are different from the traditional ESCON connections that are familiar to most S/390 owners. You can use an ESCON cable with an MT-RJ connector on one end (for the channel connection) and a traditional ESCON connector on the other end (for the control unit). Or you can use conversion cables to adapt existing ESCON cables for use with the MT-RJ connectors.

ESCON channels do not support channel spanning, and therefore cannot be shared across logical channel subsystems.

Consideration for ESCON conversion channels

There are considerations for ESCON conversion channels (which are connected to an ESCON convertor). When one of the conversion channel types (CHPID type CVC, or CBY) is defined, the channel hardware expects that an ESCON convertor is connected to the channel. If the convertor is not connected, a permanent hardware error may be reported at POR. We recommend that you do not define a conversion channel type until the convertor is actually connected.

ESCON channels with CHPID type CVC or CBY cannot be shared across logical partitions. They can only be defined as a reconfigurable channel.

5.3.3 FICON Express2 planning

FICON Express and FICON Express2 may be carried forward from previous zSeries servers when upgrading. Only FICON Express2 can be ordered (as new component) with a z9-109 system. The FICON Express2 LX feature does not support FICON Bridge (CHPID type FCV).

FICON channels can also work as Fiber Channel Protocol (FCP) channels to access SCSI devices on a Storage Area Network, for example. There is no orderable feature for FCP channels. Any FICON can work as an FCP channel by specifying the CHPID type as FCP.

FCP point-to-point attachments are now supported with FICON Express and FICON Express2 features. Point-to-point connections may be used to access data stored on devices without the use of a Fibre Channel switch. Also, an operating system or other standalone program can be IPLed via a point-to-point connection, using the SCSI IPL feature.

N_Port ID Virtualization (NPIV) provides improved sharing and channel utilization of FCP channels among operating system images in LPARs or virtual machines with the z9. To utilize NPIV, the Fibre Channel switch to which the FCP channel directly attaches is required to

support NPIV. Note that N_Port ID Virtualization is not supported with FCP point-to-point attachments.

5.3.4 OSA-Express2 planning

z9-109 systems support OSA-Express and OSA-Express2 features. Only OSA-Express2 features may be ordered as new features. OSA-Express features can be brought forward when upgrading an older machine. In the following discussion any reference to OSA-Express2 also applies to OSA-Express unless explicitly stated otherwise.

Some OSA features are not supported on the z9-109 (and were not supported on z990 and z890); these include OSA-2, OSA-2 FDDI, and OSA-Express 155 ATM. Furthermore, OSA-Express Token Ring is not supported on z9-109 systems. Possible methods of replacing these functions are:

- ▶ OSA-2: Can be replaced by the equivalent OSA-Express2 feature. A new connector may be required. All LAN Channel Stations (LCS) functions are still available, that is, you can use the OSA-Express2 feature in non-QDIO mode and have TCP/IP passthru and SNA on the same port.
- ▶ OSA-2 FDDI: Can be replaced by OSA-Express2 and a multi-protocol switch or router. The OSA-Express2 would have an Ethernet connection to the switch or router and that device would perform the FDDI conversion.
- ▶ OSA-Express 155 ATM: Can be replaced by OSA-Express2 1000BASE-T Ethernet or OSA Express2 Gigabit Ethernet and a multi-protocol switch or router with the appropriate network interface.
- ▶ OSA-Express Token Ring: Can be replaced by OSA-Express2 1000BASE-T Ethernet or OSA Express2 Gigabit Ethernet and a multi-protocol switch or router with the appropriate network interface.

The OSA-Express Integrated Console Controller function, using an OSA-Express2 1000BASE-T Ethernet feature, should be considered as a replacement for (or instead of) 3174 or 2074 control units. The OSA-ICC is a standard option within the OSA-Express2 1000BASE-T Ethernet feature and is made available through the CHPID=OSC coding on HCD.

Another OSA-Express2 function, new with z9-109 systems, is the integrated Channel Data Link Control function (OSA-CDLC) using OSA-Express2 Gigabit Ethernet and OSA-Express2 1000BASE-T Ethernet adapters. OSA-NCP (OSN) can be used in some cases to replace Communication Control Units such as IBM 372 and IBM 3745 units running NCP. This feature is standard within OSA-Express2 Gigabit Ethernet and/or 1000BASE-T Ethernet (CHPID=OSN) adapters and must be used with the appropriate Linux application running in an LPAR.

5.3.5 Coupling link connectivity

Coupling link connectivity should be planned carefully, especially if you have 9672 G5 or G6 participating in a Parallel Sysplex environment. For this type of configuration, the coupling link connectivity (ICB-2 and ISC-3 in compatibility mode) from 9672 G5 or G6 to the z9-109 server is not supported.

The connection between the z9-109 and z900 or z800 servers can be established with ICB-3 or ISC-3 links in peer mode, although ICB-4 is recommended wherever possible.

The connection with z990 or z890 servers can be done with ICB-3, ICB-4, or ISC3 links in peer mode.

The ICB-3 and ICB-4 links are restricted by the distance between servers to be connected. They cannot be more than 7 meters apart.

5.4 I/O configuration definition and management

Compared to pre-z990 servers, the z9-109 system requires a new approach to initial IOCP/IOCDS setup and an awareness of the layer of I/O definitions associated with PCHIDs. Every CHPID (except internal CHPIDs) must have a PCHID associated with it.

The z9-109 system does *not* have default CHPIDs assigned to channel ports as part of the initial configuration process performed by IBM. CHPIDs are assigned to physical channel path identifiers (PCHIDs) in the IOCP input file. It is the customer's responsibility to perform these assignments by using the HCD/IOCP definitions or by importing the output of the CHPID Mapping Tool.

Multiple Subchannel Sets (MSS)

MSS is a new item to be addressed during the I/O configuration definition on z9-109 systems. You can specify a subchannel set during HCD definitions. By default (when no SCHSET operands are present in an IOCP or HCD definition) all subchannels are allocated on MSS 0, including DASD aliases. That is, the default produces an IOCP/IOCDS without MSS dependencies, which may be suitable for many users. If you intend to use MSS, recall that there are operating system dependencies and coexistence concerns.

CHPID Mapping Tool (CMT)

The CHPID Mapping Tool (available from IBM Resource Link) provides a mechanism to map CHPIDs onto PCHIDs as required on a z9-109 system, and z990 and z890 servers. Additional enhancements have been built into the CMT to cater to the new requirements of the System z9 and zSeries servers; it provides the best availability recommendations for the installed I/O features and defined configuration.

Recall that it is possible to have up to four sets of 256 channels on a z9-109 system, that is, in four logical channel subsystems (LCSSs). Each set of channels could have, for example, four CHPIDs X'10': one in each LCSS. The manner to identify each CHPID exclusively is by associating a different PCHID to each CHPID. The CHPID Mapping Tool can help with this task.

The CHPID Mapping Tool takes input from two sources:

1. The Configuration Report file (CFreport) produced by the IBM order tool (e-Config) can be obtained from Resource Link.
2. An IOCP statement file.

The following output is produced by the CMT:

- ▶ Tailored reports. All reports should be saved for reference. The Port Report sorted by CHPID number and location should be supplied to your IBM hardware service representative for the zxxx installation.
- ▶ An IOCP input file with PCHIDs mapped to CHPIDs. This IOCP input file can then be migrated back into HCD, from which a production IODF can be built.

Important: When an IOCP statement file is exported from a Validated Work IODF using HCD, it must be imported back to HCD for the process to be valid. The IOCP file cannot be used directly by the IOCP program.

The configuration management process for a new installation of a z9-109 is reflected in Figure 5-4.

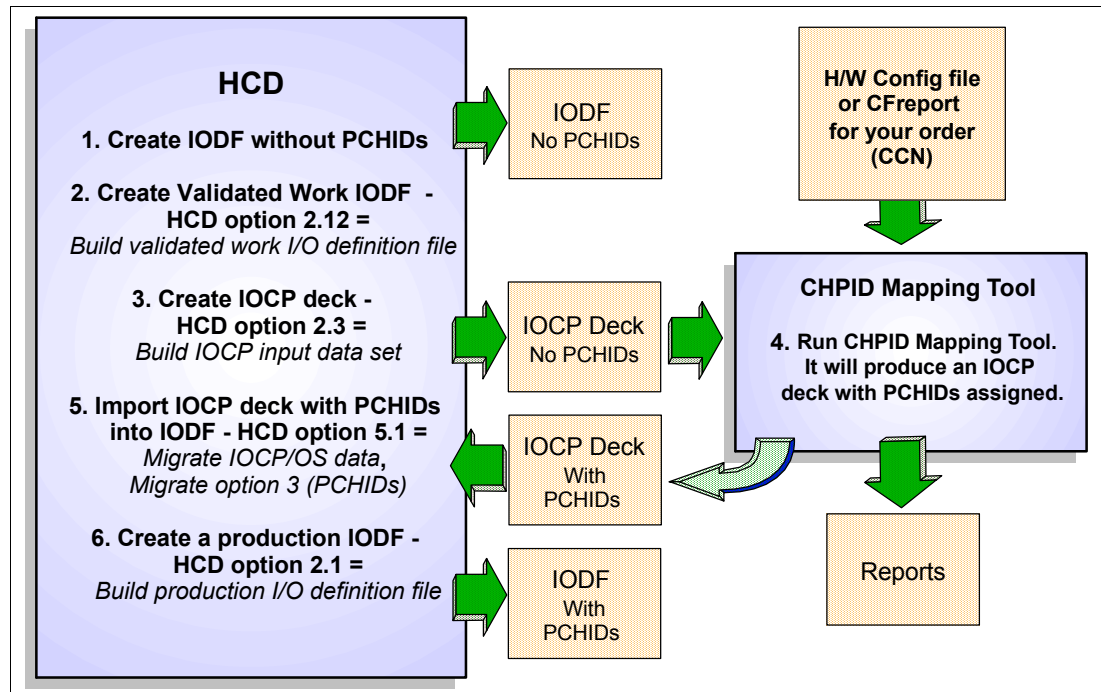


Figure 5-4 z9-109 I/O configuration definition flow

For a comprehensive explanation of the CMT, refer to *IBM eServer zSeries Connectivity Handbook*, SG24-5444.

Stand-alone IOCP

Stand-alone IOCP programs now run in a logical partition. On earlier machines, IOCP stand-alone required basic mode to run. If you are moving from a z900 or previous server to a z9-109 server and you are running an earlier operating system version not supported on a z9-109 server, you may have to create a stand-alone IOCP to start up the new environment. Note that the z9-109 Support Elements do not have diskette drives. To work directly with the IOCP source, you must import it from the HMC diskette drive or FTP it from another system on the LAN.

For more information on IOCP definitions, refer to *Input/Output Configuration Program User's Guide*, SB10-7037.

5.5 Capacity upgrades

This section describes the capacity upgrades that are offered with the z9-109 server. The capacity upgrades can be either permanent or temporary:

- ▶ Permanent
 - Ordering process (MES)
 - Capacity Upgrade on Demand (CUoD)
 - Customer Initiated Upgrade (CIU)
- ▶ Temporary
 - On/Off Capacity on Demand (On/Off CoD)

- Capacity Backup (CBU)

While capacity upgrades to the z9-109 server are concurrent, your software may not be able to take advantage of the increased capacity without performing an Initial Programming Load (IPL) or an image activation.

5.5.1 Capacity Upgrade on Demand (CUoD)

Capacity Upgrade on Demand (CUoD) allows for the non-disruptive addition of Central Processors (CPs), Internal Coupling Facilities (ICFs), Integrated Facilities for Linux (IFLs), and zSeries Application Assist Processors (zAAPs). CUoD can quickly add capacity up to the maximum number of available inactive PUs. The CUoD function, combined with Parallel Sysplex technology, enables virtually unlimited capacity upgrade options.

The CUoD functions are:

- ▶ Non-disruptive CP, ICF, IFL, and zAAP upgrades within minutes
- ▶ Dynamic upgrade of all I/O cards in the I/O cage
- ▶ Dynamic upgrade of spare installed memory
- ▶ Plan Ahead and Concurrent Conditioning

Plan Ahead and Concurrent Conditioning

Concurrent Conditioning (FC 1999) configures a system for hot plugging of I/O based on a future specified target configuration. Concurrent Conditioning of the z9-109 I/O is minimized by the fact that all I/O cards plugging into the z9-109 I/O cage are hot pluggable. This means that the only I/O to be conditioned is the I/O cage itself. The question of whether to concurrently condition an I/O cage is an important consideration, especially with the rapid change in the IT environment as well as the technology. Migration to FICON Express2 or additional OSA-Express2 networking is exceptionally easy and non-disruptive with the appropriate LIC, and if the I/O cage space is available.

The z9-109 server supports concurrent memory upgrade. This capability allows a system's memory to be increased without disrupting operation, if additional memory is physically present in an S08 or if you have a multi-book configuration.

The Plan Ahead (FC 1995) process can easily identify the configuration that is required to meet your future needs. The result of concurrent conditioning is a flexible IT infrastructure that can accommodate unpredictable growth in a low risk, non-disruptive way.

5.5.2 Customer Initiated Upgrade (CIU)

Customer Initiated Upgrade (CIU) is designed to allow you to respond to sudden increased capacity requirements by downloading and applying a Processor Unit (PU) and/or memory upgrade via IBM Resource Link and the Remote Support Facility. With the Express option on CIU, an upgrade may be installed within a few hours after order submission.

Orders (MESs) of Processor Units (PUs) and memory can be delivered by Licensed Internal Code Configuration Control (LIC CC), which is eligible for CIU (FC 9898) delivery. This includes the upgrade of PUs and memory upgrades, up to the maximum available on the installed system.

5.5.3 On/Off Capacity on Demand (On/Off CoD)

On/Off Capacity on Demand (On/Off CoD) is used when you need short-term additional capacity. On/Off CoD is designed to temporarily turn on previously uncharacterized Processor

Units (PUs), or any unassigned IFLs that are available within the current machine; as CPs, ICFs, IFLs, and zAAPs.

The maximum number of On/Off CoD zAAPs available for z9-109 servers cannot exceed the total number of zAAPs in the z9-109 server. That is, the sum of zAAPs and On/Off CoD zAAPs cannot exceed the total number of CPs.

Activation of this capability is mutually exclusive with Capacity Backup Upgrade (CBU) activation. Both On/Off CoD and CBU can reside on the z9-109 server, but only one can be activated at a time. On/Off CoD is delivered through the function of Customer Initiated Upgrade (CIU). To participate in this offering, you must have CIU Enablement (FC 9898) and On/Off CoD Enablement (FC 9896) installed. Subsequently, you may concurrently install temporary capacity, by ordering the following:

- ▶ On/Off CoD Active CP-Day (FC 9897)
- ▶ On/Off CoD Active IFL-Day (FC 9888)
- ▶ On/Off CoD Active ICF-Day (FC 9889)
- ▶ On/Off CoD Active zAAP-Day (FC 9893)

You will be billed for each On/Off CoD capacity turned on in any given 24-hour period continuing until such On/Off CoD capacity is turned off. Each month your bill is calculated for the sum of all orders installed in the prior month. Monitoring occurs through the server call home facility and a bill is generated if the capacity has been enabled for any portion of a calendar month. You will continue to be billed for use of temporary capacity until you return the server to the original state. After returning to the original state, you may choose to activate a new On/Off CoD upgrade, which can be different from the previous upgrade. When you dispose of the z9-109 server, or decide that you want to disable future temporary upgrades, you are required to remove the enablement feature, On/Off CoD Enablement (FC 9896).

One no-charge test period per contract is available for a maximum of 24 hours, starting with the activation of an On/Off CoD order. Tests that exceed 24 hours are treated in their entirety as billable On/Off CoD upgrades.

5.5.4 Capacity Backup (CBU)

The z9-109 server Capacity Backup Upgrade (CBU) capability is a non-disruptive temporary addition of CPs, IFLs, ICF, and zAAPs. With the CBU feature installed, enterprises can provide flexible, cost-effective Disaster Recovery. It makes it possible to concurrently increment the capacity of your system, using LIC CC, in the event of an unforeseen loss of substantial System z9 and zSeries computing capability at one or more of your eligible sites.

A Special Bid Contract for CBU must be approved before the CBU features can be ordered. The feature identifies how many CPs, IFLs, ICF, and zAAPs are in the Capacity Backup to IBM's vital product data base for the customer system. This contract is required and the model must be configured for the system when additional capacity is invoked. Temporary upgrade, test, or downgrade is executed via the HMC.

The CBU features are as follows:

- ▶ FC 7820 (for CPs)
- ▶ FC 7821 (for IFLs)
- ▶ FC 7822 (for ICFs)
- ▶ FC 7824 (for zAAPs)

5.6 Logical planning

In this segment we cover some considerations regarding the *logical planning* for the new z9-109 servers. Concerns such as LPAR definitions, coexistence on a Parallel Sysplex with other zSeries servers and z/OS operating systems, migration from a previous operating system version to a new version that contemplates the new z9-109 features, will be addressed. We will assume that:

- ▶ The oldest hardware you are migrating from, or replacing, is a z900 server.
- ▶ The oldest z/OS operating system you are running on a single image and/or sysplex environment is z/OS V1 R4.
- ▶ z/OS V1 R4 is running with the z990 exploitation feature installed.

If these assumptions are not valid for you, you should ask your IBM zSeries Field Technical Support to help you plan the z9-109 server migration or installation. In this case, it is possible that you will have intermediate steps to go through before moving to the zxxx server that are not covered here.

5.6.1 LPAR considerations

A z9-109 system can have up to 60 logical partitions and must have at least one LPAR because basic mode (without LPARs) is not available. The logical partition ID is a 1-byte field. Older software may work with 4-bit fields, which could be a concern when creating a system with more than 15 LPARs. The logical partition ID is specified by the user as part of the LPAR activation profile on the HMC. The following points may be important when planning LPAR definitions:

- ▶ A maximum of 15 logical partitions may be associated with a single logical channel subsystem.
- ▶ Operating systems with *compatibility support*²⁷ can only be activated in a logical partition related to LCSS 0 with a logical partition ID equal or smaller than 15 (x'F').
- ▶ Partition identifiers, names, and numbers associated with a logical partition:
 - *Logical partition identifier*. This is an identifier in the range 0 - X'3F' and is assigned by the user when defining logical partition image profiles through the SE or HMC. It is unique across the z9-109 system. This identifier is returned by an STIDP or STSI instruction. It is also known as the user logical partition ID (UPID).
 - *MIF Image identifier*.²⁸ This identifier is defined through HCD or IOCP and is the partition number defined in the RESOURCE statement in the IOCP source. It is in the range x'1' to x'F' and is unique *within* an LCSS. It does not need to be unique within a z9-109 system. The MIF Image identifier is also known as the IID.
 - *Logical partition name*. This name is defined through HCD or an IOCP and is the partition name in the RESOURCE statement in the IOCP source. A logical partition name must be unique across the z9-109 system.

Figure 5-5 on page 75 shows where the LCSS, logical partition, and MIF definitions are defined.

²⁷ Compatibility support is related to pre z/OS V1 R4 operating systems. This feature allows the recognition of a LPAR with more than x'F' for logical partition identifier and logical partition number.

²⁸ MIF is Multiple Image Facility. It was formerly known as EMIF for ESCON Multiple Image Facility when introduced for ESCON channels. This function now relates to all I/O, so the "E" has been dropped.

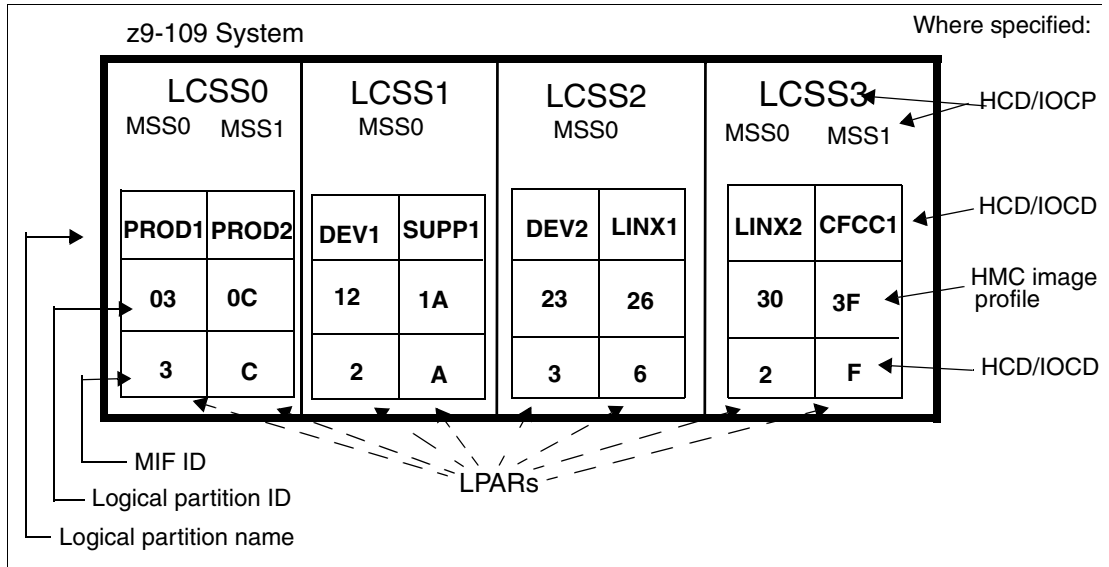


Figure 5-5 LPAR name definitions

The logical partition ID is specified when defining activation profiles. The HCD field for the MIF Image ID has the label “Partition Number.” This is equivalent to the MIF Image ID used in the IOCP RESOURCE statement. We suggest you establish a naming convention for the logical partition identifiers. For example, you could use the LCSS number concatenated to the MIF Image ID, which means logical partition ID 1A is in CSS 1 with MIF ID A. This fits within the allowed range of logical partition IDs and conveys useful information to the user.

Dynamic add/delete of a logical partition name

The ability to add meaningful logical partition names to the configuration, without a Power-On Reset, is now available. Prior to this support, extra logical partitions were defined by adding reserved names in the Input/Output Configuration Data Set (IOCDs), but it was often difficult to assign meaningful names in advance.

Dynamic add/delete of a logical partition name allows reserved logical partition “slots” to be created in an IOCDs in the form of extra logical channel subsystem (LCSS), Multiple Image Facility (MIF) image ID pairs. A reserved partition is defined with the partition name placeholder * (asterisk), and cannot be assigned to access a candidate list of channel paths or devices.

These extra logical channel subsystem MIF image ID pairs (LCSSID/MIFID) can later be assigned a logical partition name for use (or later removed) via dynamic I/O commands using HCD. The IOCDs still must have the extra I/O slots defined in advance since many structures are built based upon these major I/O control blocks in the Hardware System Area (HSA). This dynamic LPAR support is available for the z9-109 system and the z990 and z890 servers and is applicable to z/OS V1.6, and following releases.

When a logical partition is renamed, its name can be changed from NAME1 to * (with a dynamic activate) and then changed again from * to NAME2 (with another dynamic activate). The logical partition number and MIFID are retained across the logical partition name change. However, the master keys in Crypto Express2 card that were associated with the old logical partition NAME1 are retained. There is no explicit action taken against a cryptographic component for this.

5.7 Coupling Facility planning

The z9-109 system, as the z890 and z990 servers, do not provide a special model for a CF-only processor (like the model 100 for z900 processors). You can, however, have a z9-109, z990 or z890 processor with all enabled PUs defined as ICFs. Some precautions should be taken related to the following:

- ▶ The differences between z9-109 coupling link connectors and zSeries model connectors
- ▶ The ICB link distances and their incompatibility with earlier servers

Coupling Facility Level 13 (CFLEVEL 13) introduced the application of patches to Coupling Facility Control Code (CFCC) concurrently, in “rolling” mode. As shown in Figure 5-6, on a single CEC you can first apply the patch on a test ICF LPAR, perform all necessary tests, and later on roll the patch to the production logical partition. To use the updated CFCC code in a CF logical partition, simply deactivate and activate the partition. When the CF comes up, it displays its version on the OPRMSG panel for that partition.

With this facility there is no need to stop the entire machine for CFCC patches; however, CFLevel upgrades will still be disruptive to the entire CEC.

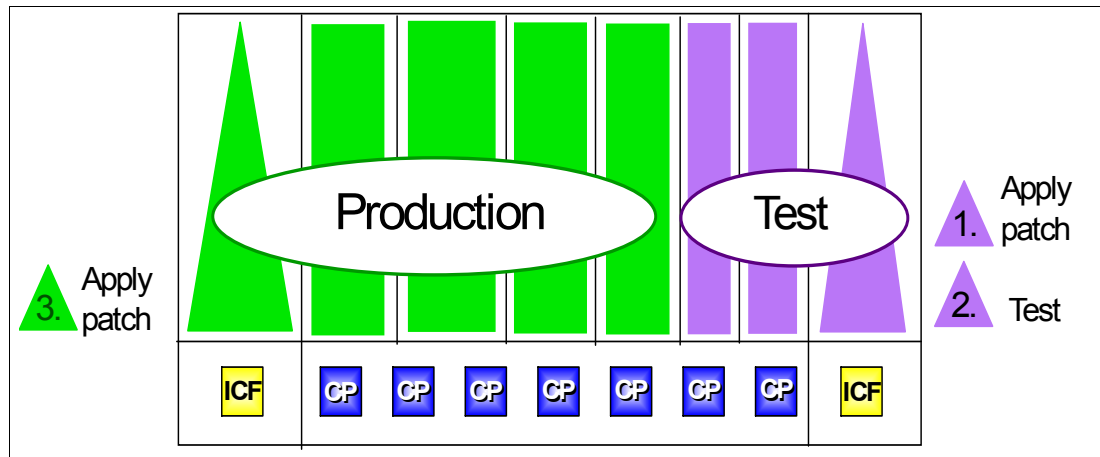


Figure 5-6 Concurrent CFCC patch

5.7.1 Sysplex Timer ETR Network ID

As part of the installation of a Sysplex Timer in either Basic or Expanded Availability configuration, each IBM 9037 Sysplex Timer Unit is assigned an ETR Network ID (0 to 31decimal) and ETR Unit ID (0 to 31decimal). The ETR network ID and ETR Unit ID values are arbitrary (within the valid range) and can be chosen by the customer to uniquely identify an ETR network and a unique ETR unit (Sysplex Timer) within the ETR Network.

For example, in a Sysplex Timer Expanded Availability configuration there are two IBM 9037 Sysplex Timer Units. The first IBM 9037 may have a ETR Network ID value of 0, and a ETR Unit ID value of 0. The second IBM 9037 will have the same ETR Network ID value of 0, but a different ETR Unit ID value of 1. The two Sysplex Timer Units in an Expanded Availability configuration must have the same ETR Network ID value defined. However, their ETR Unit ID values must be unique within the Sysplex Timer ETR Network.

A Sysplex Timer Unit's ETR Network ID, ETR Unit ID, and Port number are transmitted along with timing signals to an attached server's ETR port. This information is available to sysplex systems running on the server and can be displayed using the D ETR command.

A function, introduced with the z990 server, and present in the z9-109 system, requires the ETR Network ID of the attached Sysplex Timer Network to be manually set in the Support Element at installation time. In addition, on the same SE panel, the ETR ports must be enabled for stepping the TOD. This function checks whether the ETR Network ID being received in the timing signals via each of the server's two ETR ports matches the ETR Network ID manually set in the server's Support Element (SE).

This function provides greater checking capability, helping to eliminate cabling errors where a server's ETR port may be incorrectly connected to a Sysplex Timer Unit in an incorrect Sysplex Timer ETR Network, and allows verification of cabling connectivity from the Sysplex Timer to the z9-109 system prior to IPL of z/OS.

Important: Setting the ETR Network ID to an incorrect value on a z9-109 system's SE will cause all operating systems with ETRMODE=YES on the server to immediately enter a non-restartable disabled wait state.

Ordering ETR cables

Two ESCON MT-RJ cables for the ETR feature on the zSeries zxxx system will automatically be added to any zSeries Fiber Cabling Service contract offered by IBM Global Services. Contact your local IBM Installation Planning Representative, IBM z9-109 system Product Specialist, or IBM Connectivity Services Specialist for details. If you choose not to use this service, you may purchase the two ESCON MT-RJ cables separately, or provide them yourself from another source. An ESCON MTRJ-to-ESCON Duplex conversion kit is also needed to connect the MT-RJ cables to the Duplex connectors on the 9037.

Sysplex Timer attachment planning

ETR ports on z9-109 systems are optional features and must be ordered separately. However, if you order any CF links, an order for ETR ports is generated automatically.

5.8 Operating system migration

Operating system support for z9-109 systems is summarized in Table 5-1. If you are running any earlier operating system not listed in Table 5-1, you will need to migrate to a more recent operating system version.

Table 5-1 Operating system service dates

Operating System	z9-109	z990 z890	z900 z800	Ship Date	End of Service	Coexist
z/OS V1 R4 ^a	yes	yes	yes	09/02	03/07	z/OS V1R7
z/OS V1 R5 ^b	yes	yes	yes	03/04	03/07	z/OS V1R7
z/OS V1 R6	yes	yes	yes	09/04	09/07	
z/OS V1 R7	yes	yes	yes	09/05	09/08	
z/VM V4 R4 ^c	yes	yes	yes	08/03	09/06	
z/VM V5 R1	yes	yes	yes	09/04	09/07	
z/VM V5 R2	yes	yes	yes	1Q06		
VSE/ESA V2.7 ^d	yes	yes	yes	03/03		
z/VSE V3 R1 ^d	yes	yes	yes	03/05		

Operating System	z9-109	z990 z890	z900 z800	Ship Date	End of Service	Coexist
TFP V4 R1	yes	yes	yes	02/01		
z/TFP V1 R1	yes	yes	yes	09/05		

- a. z/OS V1 R4 is no longer orderable. The z990 exploitation feature is orderable until 12/06.
- b. No longer orderable.
- c. Can run either in 31- or 64-bit mode.
- d. Runs only in 31-bit mode and does not implement zArchitecture.

Stand-alone z9-109 systems can run any of the operating systems above with the PSP-recommended service applied. However, exploitation of the z9-109 hardware features depends on the operating system version used.



A

Frequently asked questions

Q: Can the MIDA W list reside above 2 GB?

A: No. However the target of the MIDA W (the data address) can be above 2 GB.

Q: Does the new HMC contain 3270 sessions? Can they be used as a full-function MVS™ console?

A: The 3270 sessions are available but cannot be used as MVS operator consoles.

Q: Can the HMC 3270 session be used as a TSO terminal?

A: No, it is not supported by VTAM.

Q: Will z/OS 1.7 run on MP3000 systems?

A: No. z/OS 1.7 requires Architecture Level Set 3, commonly known as “64-bit operation.”

Q: Is a raised floor required for the z9-109 system?

A: For practical purposes, yes. The cables could be handled above the floor but the heat load would be very difficult to handle without a raised floor and an appropriate chilled air flow.

Q: Can a normal application program use MIDA Ws in a channel program?

A: Not directly. Most “normal” application programs that create their own channel programs use the EXCP interface. This does not support the use of MIDA Ws. An application program can use MIDA Ws via access methods that use the media manager interface, but this does not allow programers to build their own CCWs and MIDA Ws.

Q: Do MIDA Ws replace all data chaining CCWs?

A: No. MIDA W usage is limited to DASD. Also, note that the total MIDA W data count (in a MIDA W list) cannot exceed 64K bytes. This is sufficient for any CKD DASD, but may not be sufficient in all cases for other devices, such as tape drives.

Q: Can a CCW that uses MIDA W also use the data chaining bit to use the next CCW in a data chaining mode?

A: No.

Q: Do we still need the processor affinity controls to use the cryptographic instructions?

A: No.

Q: Are the cryptographic instructions (in every PU) always enabled?

A: No. They must be enabled via an SE function. (The hashing functions provided through these instructions are always available.)

Q: Does the Crypto Express2 adapter work with public keys (RSA) longer than 512 bits?

A: All public functions of the Crypto Express2 adapter are designed to work with keys *longer* than 511 bits. In addition, short RSA clear-key operations (less than 512 bits) are supported for the following functions (using the ICSF names): Digital Signature Verify, Public Key Encrypt, Public Key Decrypt.

Q: Is Expanded Storage supported on the z9-109 system?

A: Yes, it is supported by the hardware. However, it is not used by current z/OS versions. It is used by z/VM.

Q: What are the real differences between the FICON Express and FICON Express2 adapters? I thought that FICON was a standard protocol; why are different adapters needed?

A: The differences are in the speeds supported and the number of ports (CHPIDs) per adapter. FICON is a standard protocol that allows more than one link speed.

Q: I was told that assembly language programming is bad for the zSeries processors and probably for the z9-109 system. Is this true?

A: No, that is not correct. The performance of these processors has been enhanced by the use of superscalar operation with multiple internal pipelines that allow several instructions to be in process at the same time. Also, cache memory operates most efficiently when its alignment is taken into consideration and L1 cache overlaps are avoided. In principle, all this is transparent to the program (and the programmer). However, peak performance depends on an instruction stream designed to best exploit the superscalar, pipeline, and cache functions.

Recent compilers may automatically order the instruction stream to best comply with this design. Assembly programs written in a logical order (as seen by the programmer) might not have instructions in the best order for optimal performance. The programs continue to work, of course, and the effects can be ignored for the vast majority of assembly programs. In a few cases involving very heavily-used code, reexamining the code might produce a measureable improvement in performance. (In principle, the same considerations could apply to programs compiled by older compilers.)

Q: Can the new instructions be used from assembler? What compilers will use the new instructions? When?

A: Yes, the new instructions can be used in an assembly language program via a PTF for the assembler. We do not know when various compilers will start using the new instructions.

Q: Can the program-directed re-IPL be used for z/OS? How?

A: We assume it could be used since it functions with ECKD (CCW) IPLs. We have no programming information at the time of writing.

Q: What happens (with the *wild branch* function) if the target code performs several successful branches before it ABENDs?

A: The *branch from* information is not useful in this case. It records only the most recent successful branch instruction address.



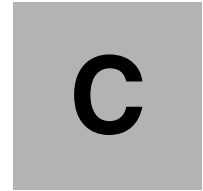
Instruction set changes

The following is a listing of the new and changed instructions for the z9-109 system. This list is intended only to provide a general flavor of the instructions. More specific details will be available in a future Principles of Operation manual.

Note	Op	Mn	Instruction name
	B338	MYR	Multiply unnormalized (ExtHFP <--LongHFP x LongHFP)
	B33D	MYHR	Multiply unnormalized (ExtHFP[high] <--LongHFP x LongHFP)
	B339	MYLR	Multiply unnormalized (ExtHFP[low] <-- LongHFP x LongHFP)
	ED3B	MY	Multiply unnormalized (ExtHFP <-- LongHFP x LongHFP)
	ED3D	MYH	Multiply unnormalized (ExtHFP[high] <--LongHFP x LongHFP)
	ED39	MYL	Multiply unnormalized (ExtHFP[low] <--LongHFP x LongHFP)
	B33A	MAYR	Multiply & add unnormalized (ExtHFP <-- LongHFP sources)
	B33C	MAYHR	Multiply & add unnormalized (ExtHFP[high] <-- LongHFP sources)
	B338	MAYLR	Multiply & add unnormalized (ExtHFP[low] <-- LongHFP sources)
	ED3A	MAY	Multiply & add unnormalized (ExtHFP <-- LongHFP sources)
	ED3C	MAYH	Multiply & add unnormalized (ExtHFP[high] <-- LongHFP sources)
	ED38	MAYL	Multiply & add unnormalized (ExtHFP[low] <-- LongHFP sources)
	C29	AFI	Add immediate (32)
	C28	AGFI	Add immediate (64 <-- 32)
	C2B	ALFI	Add logical immediate (64 <-- 32)
	COA	NIHF	Add immediate [high]
	COB	NILF	Add immediate [low]
	C2D	CFI	Compare immediate (32)
	C2C	CGFI	Compare immediate (64 <-- 32)
	C2F	CLFI	Compare logical immediate (32)
	C2E	CLGFI	Compare logical immediate (64 <-- 32)
	C06	XIHF	Exclusive OR immediate [high]
	C07	XILF	Exclusive OR immediate [low]
	B983	FLOGR	Find leftmost one
	C08	IIHF	Insert immediate [high]
	C09	IILF	Insert immediate [low]
	E312	LT	Load and test (32)
	E302	LTG	Load and test (64)
	B926	LBR	Load byte (32)
	B906	LGBR	Load byte (64)
	B927	LHR	Load halfword (32)
	B907	LGHR	Load halfword (64)
	C01	LGFI	Load immediate (64 <-- 32)

	E394	LLC	Load logical character (32)
	B994	LLCR	Load logical character (32)
	B984	LLGCR	Load logical character (64)
	E395	LLH	Load logical halfword (32)
	B995	LLHR	Load logical halfword (32)
	B985	LLGHR	Load logical halfword (64)
	C0E	LLIHF	Load logical immediate [high]
	C0F	LLILF	Load logical immedaite [low]
	C0C	OIHF	OR immediate [high]
	C0D	OILF	OR immediate [low]
	C25	SLFI	Subtract logical immediate (32)
	C24	SLGFI	Subtract logical immediate (64 <-- 32)
a	B92E	KM	Cipher message
a	B92F	KMC	Cipher message with chaining
a	B93E	KIMD	Compute intermediate message digest
a	B93F	KLMD	Compute last message digest
	B9AA	LPTEA	Load page-table-entry address
	B27C	STCKF	Store clock fast
ab	B993	TR00	Translate one to one
ab	B992	TR0T	Translate one to two
ab	B991	TRT0	Translate two to one
ab	B990	TRTT	Translate two to two
	B2B0	STFLE	Store facility list extended
ab	B9B1	CU24	Convert UTF-16 to UTF-32
ab	B2A6	CU21	Convert UTF-16 to UTF-8
ab	B9B3	CU42	Convert UTF-32 to UTF-16
ab	B9B2	CU41	Convert UTF-32 to UTF-8

a- new function(s) added to existing instruction
b- RRE-format instruction changed to RRF-format



Channel options

Table C-1 lists the channel options supported on the z9-109 server with required connector and cable types, and the maximum unrepeated distances, as well as the maximum number of features and connections.

Note that a model S08 supports a maximum of 64 I/O features, and at least one I/O feature is required with all z9-109 system models.

Table C-1 z9-109 channel feature support

Channel feature	Feature codes	Maximum		Connector type	Cable type	Maximum unrepeated distance
		Features	Connections			
16-port ESCON	2323 ^a	69	1024 channels	MT-RJ	MM 62.5 μm	3 km
FICON Express LX ^b	2319	60	120 channels	LC Duplex	SM 9 μm	10 km/20 km ^c
FICON Express SX ^b	2320	60	120 channels	LC Duplex	MM 62.5 μm MM 50 μm	120 m 300 m
FICON Express2 LX	3319	84	336 channels	LC Duplex	SM 9 μm	10 km/12 km ^c /20 km ^c
FICON Express2 SX	3320	84	336 channels	LC Duplex	MM 62.5 μm MM 50 μm	120 m 300 m
OSA-Express GbE LX ^b	1364	24	48 ports	LC Duplex	SM 9 μm	5 km
					MCP ^g	550 m
OSA-Express GbE LX ^b	2364	24	48 ports	SC Duplex	SM 9 μm	5 km
					MCP	550 m
OSA-Express GbE SX ^b	1365	24	48 ports	LC Duplex	MM 62.5 μm	220 m 275 m
					MM 50 μm	550 m

Channel feature	Feature codes	Maximum		Connector type	Cable type	Maximum unrepeat distance
		Features	Connections			
OSA-Express GbE SX ^b	2365	24	48 ports	SC Duplex	MM 62.5 µm	220 m 275 m
					MM 50 µm	550 m
OSA-Express 1000BASE-T Ethernet ^b	1366	24	48 ports	RJ 45	UTP Cat 5	100 m
OSA-Express Fast Ethernet ^b	2366	24	48 ports	RJ 45	UTP Cat 5	100 m
OSA-Express2 GbE LX	3364	24	48 ports	LC Duplex	SM 9 µm	5 km
					MCP	550 m
OSA Express2 GbE SX	3365	24	48 ports	LC Duplex	MM 62.5 µm	220 m 275 m
					MM 50 µm	550 m
OSA-Express2 1000BASE-T Ethernet	3366	24	48 ports	RJ 45	UTP Cat 5	100 m
OSA-Express2 10 GbE LR	3368	24	24 ports	SC Duplex	SM 9 µm	10 km
HiperSockets	n/a	n/a	16 internal LANs	n/a	n/a	n/a
IC	n/a	n/a	32 links	n/a	n/a	n/a
ICB-3 ^d	0993	16	16 links	FC 0227		10 m
ICB-4 ^d	3393	16	16 links	FC 0228		10 m
ISC-3 (peer mode) ^e	0217 ^e	12	48 links	LC Duplex	SM 9 µm	10 km/12 km ^c /20 km ^c
ISC-3 (RPQ 8P2197 / Peer mode at 1 Gbps) ^f	0218 0219					
Crypto Express2 ^g	0863	8	16 PCI-X adapters	n/a	n/a	n/a
ETR link	6154	2	2 ports	MT-RJ	MM 62.5 µm	3 km/26 km ^c
					MM 50 µm	2 km /24 km ^c

a. ESCON channels are ordered in increments of four using feature code 2324.

b. Only supported when carried forward on an upgrade.

c. Requires RPQ 8P2263. Maximum distances vary based on the link data rate.

d. The maximum number of coupling links (excluding IC) that can be ordered is 64. The total number of coupling links that can be defined is 64, including IC links.

e. There are two feature codes for the ISC-3 card:

Feature code 0217 is for the ISC Mother card (ISC-M).

Feature code 0218 is for the ISC Daughter card (ISC-D).

Individual ISC-3 port activation must be ordered using feature code 0219.

f. RPQ 8P2197 enables the ordering of a different daughter card supporting 20 km unrepeat distance for 1 Gbps peer mode. RPQ 8P2262 is a requirement for that option, and other than the normal mode the channel increment is two, that is both ports (FC 0219) at the card have to be activated.

g. The initial order is two features (4 PCI-X adapters). Each PCI-X adapter can be configured (via HMC/SE) as either a secure coprocessor or accelerator.

Related publications

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

IBM Redbooks

For information on ordering these publications, see “How to get IBM Redbooks” on page 85. Note that some of the documents referenced here may be available in softcopy only.

- ▶ *IBM eserver zSeries 990 Technical Introduction*, SG24-6863
- ▶ *IBM eserver zSeries 890 Technical Introduction*, SG24-6310
- ▶ *IBM eserver zSeries Connectivity Handbook*, SG24-5444
- ▶ *IBM eserver zSeries 990 Technical Guide*, SG24-6947

Other publications

These publications are also relevant as further information sources:

- ▶ *z/Architecture Principles of Operation*, SA22-7832

Online resources

These Web sites and URLs are also relevant as further information sources:

- ▶ IBM general Web site
<http://www.ibm.com>
- ▶ Preventive Service Planning
http://techsupport.services.ibm.com/390/psp_main.html

How to get IBM Redbooks

You can search for, view, or download Redbooks, Redpapers, Hints and Tips, draft publications and Additional materials, as well as order hardcopy Redbooks or CD-ROMs, at this Web site:

ibm.com/redbooks

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services

Index

Numerics

1000BASE-T Ethernet 56
16-port ESCON 17
3270 and ASCII consoles 62
60 logical partitions 53
63.75K subchannels 54

A

Advanced Encryption Standard operations (AES) 20
air-cooled 26
assembly language programming 80
ATM 7, 69

B

basic mode 45, 74
BM Networking Services 65
books, jumper 11
books, system 10
branch detection 42
branch history table 14
bus and tag 67

C

cable-related services 65
cabling migration 66
cache, L2 11
Capacity Backup (CBU) 72–73
Capacity Upgrade on Demand (CUoD) 71–72
CBU 72–73
CCF cryptographic coprocessors 7
ceramic carrier 12
CFCC 76
Channel Data Link Control 69
channel subsystem 30
characterizations, PU 14
CHPID Mapping Tool 44, 70
CHPID number 32
CHPID numbers 30–31
CHPID types 17
CIU 71–72
 enablement 73
clock speed 2
CMT 70
Communication Controller for Linux (CCL) 18, 37
Compatibility support feature 52
concurrent conditioning 72
Concurrent maintenance and upgrades 24
Cooling 25
Coupling connections 20
Coupling Facility (CF) 76
Coupling Facility Control Code 76
Coupling Links connectivity 69

CP 72–73
CP Assist for Cryptographic Function (CPACF) 20
CP characterization 14
CPACF (CP Assist for Cryptographic Function) 20
CPACF Enhancements 58
Crypto Express2 17, 21, 80
cryptographic functions 20
cryptographic hardware 3
cryptographic instructions 42, 80
CUoD 71–72
Customer Initiated Upgrade (CIU) 71–72

D

data chaining, CCW 39
DCAs 25
domain in an I/O cage 15
dual-core processors 13
Dynamic Address Translation (DAT) 43

E

enhanced book availability 28
enhanced driver maintenance 28
ES conversion channels 34, 68
ESAME Architecture 3
ESCON
 channel 68
ESCON channel planning 68
Ethernet adapters, HMC 63
Ethernet hub 63
ETR attachment 21
evolution, system 2
EXCP 41
EXCPVR IOS driver 41
Expanded Storage 80
Expanded Storage (ES) 52
Exploitation support feature 52
extended format data sets 39
Extended Translate Facility-2 42
Extended Translate Facility-3 42

F

fanout cards, I/O 12
FCP (Fiber Channel Protocol) 68
FCP channels 41
FCP point-to-point 68
FDDI 7, 69
fiber cable connectors 66
fiber optic cables 65
FICON Bridge (FCV) 68
FICON Express 17
FICON Express2 planning 68
FICON Express2.0 17
FICON link incident report 55

FICON link recovery thresholds 28
FICON link status 28
flexible memory option 15, 28
Flexible Support Processor (FSP) 23
Floating point operation 43

G

GARP VLAN Registration Protocol (GVRP) 38, 56

H

Hardware Management Console (HMC) 23, 45, 62
Hardware Storage Area (HSA) 22
Hardware System Area (HSA) 75
HiperSockets 17
HiperSockets IPv6 57
HMC (Hardware Management Console) 23
HMC 3270 sessions 79
HMC and SEs 23
HMC connectivity 63
HMC style 46
HSA (Hardware System Area) 75

I

I/O
 adapter card 31
 cages 15–16
 configuration definition 70
 fanout cards 12
 features 83
 interfaces 15
I/O feature slots 15
IBM 3745 69
IBM 9034 67
IBM Power PC microprocessor 23
IBM Resource Link 44
IC channel 17
ICB-2 feature 7
ICB-3 and ICB-4 cables 65
ICB-3 links 17
ICB-4 channels 20
ICB4 connections to books 20
ICB-4 links 17
ICB4/ICB3 coupling links 15
ICF 72–73
ICF characterization 14
IEEE arithmetic 43
IFL 72–73
IFL characterization 14
Input/Output Configuration Dataset (IOCDS) 75
instructions, changed 14
Integrated Facility for Linux (IFL) 72–73
Internal Battery Feature 24
Internal Coupling Facility (ICF) 72–73
IOCDS 32, 36, 75
IOCDS parameters 19
IOCP 71
IOCP definition 36
IOCP, stand-alone 71

ISC-3 links 17

L

L1 caches 13
L2 cache 42–43
large send, TCP/IP 18
LCSS 18
 ID 75
LCSS (Logical Channel Subsystem) 18, 74
Logical Channel Subsystem (LCSS) 18, 74
logical partition
 identifier 74
 name 74
LPAR 45
 identifier 45
 mode 45, 74
 name 45
 naming and numbering 32

M

MBA (Memory Bus Adapter) 12
MCM (Multiple Chip Module) 12
media manage 41
Memory Bus Adapter (MBA) 12
memory cards 14
memory mapping 29
memory sparing 28
MIDAW facility 39, 55
MIDAW list 79
MIF Image ID 45, 74
MIF image ID 75
MIF shared channel 33
MIF spanned channel 33
MIFID 75
migrating from a z900 7
model numbers 6
model numbers, software 6
modified IDAW 39
Modular Refrigeration Unit (MRU) 25
MP3000 systems 79
MRU (Modular Refrigeration Unit) 25
multimode trunk cable 67
Multiple Chip Module (MCM) 12
Multiple Image Facility (MIF) 32, 34
Multiple Subchannel Sets (MSS) 2, 34, 54, 70
multiplex functions, I/O cage 15

N

N_Port ID virtualization (NPIV) 41, 55, 68
NetBios, HMC 64
New and changed instructions 42

O

On/Off Capacity on Demand 71–72
On/Off CoD 71–72
 active CP-day 73
 active ICF-day 73
 active IFL-day 73

- active zAAP 73
- enablement 73
- open exchanges for FCP and FICON 44
- operating system requirements 51
- Optica 67
- OS/2 62
- OSA ATM 69
- OSA-2 69
- OSA-2 FDDI 69
- OSA-Express 17
- OSA-Express2 17, 39
- OSA-Express2 OSN support 57
- OSA-Express2 planning 69
- oscillator switchover 28

P

- Parallel Access to Volumes (PAV) 34
- parallel channel migration 67
- parallel channels 7
- Parallel Sysplex 3
- partition number 45, 75
- PCHID 70
- PCHID number 19, 31
- PCHID, physical channel identifier 31
- PCICC, PCICA, and PCIXCC cryptographic features 7
- PCI-X support 58
- performance characteristics 42
- physical channel identifier (PCHID) 19
- Physical Channel Path Identifiers (PCPID) 70
- Plan Ahead 72
- power and cooling 24
- power boundary 25
- PR/SM performance 43
- Preventive Service Planning (PSP) 51
- processor chips 13
- processor unit (PU) 6, 13
- program-directed re-IPL 44, 56
- Pseudo Random Number Generation 20
- PU characterizations 14
- public keys (RSA) 80
- PUs, spare 6

Q

- QDIO 3
- QDIO usage 18

R

- raised floor 79
- real memory 43
- Redbooks Web site 85
 - Contact us x
- Redundant I/O Interconnect 28, 30
- refrigeration cooling 25
- Remote Support Facility (RSF) 64
- RESOURCE statement 37, 45, 74–75
- ring design, cache 11
- RSF (Remote Support Facility) 64

S

- SAN (Storage Area Network) 41, 68
- SAP characterization 14
- SAPs 14
- SAPs, included 6
- SCHSET operand 37
- SCSI devices 68
- SE (Support Element) 23, 45, 63
- SE connectivity 63
- Secure Hash Algorithm - 256 20
- Server Time Protocol (STP) 22
- Single system image 53
- single-core processors 13
- SNA protocols, HMC 64
- spanned channel 31–32
- spare PUs 6
- STI interfaces 15
- STI links 42
- STI ports 12
- Storage Area Network (SAN) 41, 68
- storage protect keys 15
- Store facility list extended 42
- subchannel numbers 34
- superscalar design 42
- superscalar processor 13
- Support Element (SE) 23, 45, 63
- Sysplex Timer 21
 - Time-of-Day 21
- system image 53

T

- TCP/IP stacks 18
- TKE (Trusted Key Entry) 21
- TLB (Translation Lookaside Buffer) 13
- TN3270E functions 18
- token ring 7
- Translation Lookaside Buffer (TLB) 13, 43
- tree interface 62
- Trusted Key Entry (TKE) 21

U

- uncharacterized PUs 14
- upgrades, from zSeries 6

V

- VLAN ID 38, 56

W

- Web browser, HMC 24
- weight, frames 10
- wild branch 42, 80

Z

- z/Architecture 1, 3
- z9-109 1
- z9-109 system models 6
- zAAP 72–73

zAAP characterization 14
zSeries 1
zSeries Application Assist Processor (zAAP) 72–73
zSeries comparisons 4



IBM System z9 109 Technical Introduction

(0.2" spine)
0.17" x 0.473"
90 x 249 pages



IBM System z9 109 Technical Introduction



Redbooks

Hardware description

This IBM Redbook introduces the IBM System z9 109 (z9-109), which offers a continuation of the IBM scalable mainframe servers. The z9-109 server is based on z/Architecture and provides extensions to the architecture. In addition to increased performance and expansion options, particular emphasis has been placed on the integration of the z9-109 into a complete system and software environment.

Software support

Key functions

This document provides basic information about new hardware functions and features, associated software support, and migration considerations. It is intended for systems engineers, hardware planners, and anyone else needing to understand the new z9-109 system elements and plan for their usage.

This document is not intended as an introduction to mainframes. Readers are expected to be generally familiar with current zSeries technology and terminology.

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

SG24-6669-00

ISBN 0738493813